

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 42.1481 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MMLPRFSKTVTVLLAQTT.....LTLDLVLPRGDFHFSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2051	100.0	390	4	AAY72639 Human gly
2	2051	100.0	390	5	Abb81556 Human int
3	1865.5	91.0	418	3	Aab41947 Human ORF
4	1758.5	85.7	395	5	Abb81555 Consensus
5	1729.5	84.3	395	4	AAY72640 Human gly
6	1729.5	84.3	395	5	Abb81554 Human cor
7	1729.5	84.3	395	5	Aae15438 Human dru
8	1729.5	84.3	395	7	Adi21086 Novel hum
9	1729.5	84.3	395	8	Adl61235 Human tyr
10	1536.5	74.9	395	4	AAY72638 Mouse gly
11	1536.5	74.9	395	5	Aau11275 Murine in
12	1523	74.3	418	5	Abb81557 Mouse int
13	1513.5	73.8	394	7	Adj70405 Human hea
14	1013	49.4	386	4	Aam93309 Human pol
15	1013	49.4	386	8	Adl30784 Human pro
16	1008	49.1	386	2	AAY39918 Human gly
17	1003.5	48.9	380	5	Aau11274 Human L-s
18	970	47.3	386	3	AAY79219 Human tra
19	958.5	46.7	388	2	Aay39919 Mouse gly
20	808	39.4	169	5	Abb81559 Human int
21	755	36.8	169	5	Abb81558 Human cor
22	665	32.4	483	2	AAY31656 Mouse N-a
23	659.5	32.2	530	4	Aab95367 Human pro
24	659.5	32.2	530	8	Adq18590 Human sof
25	656.5	32.0	484	2	AAY31657 Human N-a

ALIGNMENTS

RESULT 1

AAY72639
ID AAY72639 standard; protein; 390 AA.

AC AAY72639;

XX
XX
DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha).

XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW hemmolytic disease; Grave's disease; hypoparathyroidism; anaemia;
KW dermatinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW chromosome 16q23.1.

XX Homo sapiens.

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; RAD02697, RAD02698, RAD02699.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 1; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4 alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2

CC membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non

CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
CC the sulphation activity of GST. GST is also useful in inhibiting a
CC selectin mediated binding event. GST is useful in gene therapy to treat
CC disorders such as acute or chronic inflammation, systemic lupus
CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, pernicious
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation. Note: The present sequence is also shown in
CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
XX

XX Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.2e-215;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPRFSSKTVTVLLLAQTTCCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Db 1 MWLPRFSSKTVTVLLLAQTTCCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Qy 61 FSOHPDVFYLMFPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVFDAYMPSQSNLSA 120
Db 61 FSOHPDVFYLMFPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVFDAYMPSQSNLSA 120
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTRQPFSLAREACRSYSHVVLKEVRFF 180
Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTRQPFSLAREACRSYSHVVLKEVRFF 180
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Qy 241 RLIREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 241 RLIREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Qy 301 EAWIHNITHGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLLYGR 360
Db 301 EAWIHNITHGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLLYGR 360
Qy 361 PVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
Db 361 PVYSADQQRDLTLDLVLPRGPDHFSWASPD 390

RESULT 2

ABB81556
ID ABB81556 standard; protein; 390 AA.

AC ABB81556;

XX 05-SEP-2002 (first entry)

XX Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
DE Human, N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.

XX Homo sapiens.

XX US2002061562-A1.

XX 23-MAY-2002.

XX 09-AUG-2001; 2001US-00927602.

XX 11-AUG-2000; 2000US-00638211.

PR 11-AUG-2000; 2000US-0325773P.

XX (FUKU/) FUKUDA M N.

PA (AKAW/) AKAWA T O.

XX Fukuda MN, Akama TO;

XX WPI; 2002-507643/54.

XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.

XX Example 5; Fig 2A-B; 69pp; English.

XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratectomy. The present sequence
CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention

XX Sequence 390 AA;

Query Match 100.0%; Score 2051; DB 5; Length 390;

Best Local Similarity 100.0%; Pred. No. 1.2e-215;

Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPRFSSKTVTVLLLAQTTCCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Db 1 MWLPRFSSKTVTVLLLAQTTCCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Qy 61 FSOHPDVFYLMFPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVFDAYMPSQSNLSA 120
Db 61 FSOHPDVFYLMFPAWHVWTTLSQSAATLHMVARDLMRSIFLCMDVFDAYMPSQSNLSA 120
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTRQPFSLAREACRSYSHVVLKEVRFF 180
Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTRQPFSLAREACRSYSHVVLKEVRFF 180
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Qy 241 RLIREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 241 RLIREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
Qy 301 EAWIHNITHGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLLYGR 360
Db 301 EAWIHNITHGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRQVEVCAGALQLLYGR 360
Qy 361 PVYSADQQRDLTLDLVLPRGPDHFSWASPD 390
Db 361 PVYSADQQRDLTLDLVLPRGPDHFSWASPD 390

RESULT 3

AAB41947

ID AAB41947 standard; protein; 418 AA.

XX AAB41947;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;

KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US008621.
XX 31-MAR-1999; 99US-0127607P.
XX 02-APR-1999; 99US-0127636P.
XX 05-APR-1999; 99US-0127728P.
XX 30-MAR-2000; 2000US-00540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; AAC76156.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX Claim 11; Page 2599-2600; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antithyroid; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
SQ Sequence 418 AA;
Query Match 91.0%; Score 1865.5; DB 3; Length 418;
Beat Local Similarity 91.0%; Pred. No. 3.1e-195;
Matches 356; Conservative 9; Mismatches 25; Indels 1; Gaps 1;
QY 1 MWLPRFSKTVVLLAQTCLLLFIIRPGPSSPAGGEDRVHVLVLSWSRSGSFLGQL 60
DB 28 MWLPRFSKTVVLLAQTCLLLFIIRPGPSSPAGGEDRVHVLVLSWSRSGSFLGQL 87
QY 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMRSLFCLDMDFDAYMVEGPPRRQS 119

Db 88 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMRSLFCLDMDFDAYMVEGPPRRQS 147
QY 120 AFNWATSRALCSPACAPPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 179
Db 148 SLQWENSRALCSPACACDIIPQDEIIPRAHRLCLCSQPFVEVKEACRSYSHVVLKEVRF 207
QY 180 FNLQVLYPLSDPALNURIVHLVRDPRVLRREAAAGPILARDNGIVLGTNGKWEADPH 239
Db 208 FNLQSLYPLKLPDSNLHLVHLVRDPRVLRREAAAGPILARDNGIVLGTNGKWEADPH 267
QY 240 LRLIREVCRSHVRIAEATLTKPPPLGRVLRVRFEDLAREPLAEIRALYAFGLTLTPQ 299
Db 268 LRLIREVCRSHVRIAEATLTKPPPLGRVLRVRFEDLAREPLAEIRALYAFGLTLTPQ 327
QY 300 LEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPTKILRVQEVCAQALQLLGY 359
Db 328 LEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPTKILRVQEVCAQALQLLGY 387
QY 360 RPVYSADQQRDLTLDLVLRGPDHFWASPD 390
Db 388 RPVYSADQQRDLTLDLVLRGPDHFWASPD 418
RESULT 4
AAB81555
XX ABB81555 standard; protein; 395 AA.
XX AC ABB81555;
XX DT 05-SEP-2002 (first entry)
XX DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX OS Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX Key Location/Qualifiers
FH Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391

FT Misc-difference 392 /label= Ser, Lys
 FT Misc-difference 394 /label= His, Gln
 FT Misc-difference 394 /label= Arg, Glu
 FT Misc-difference 395 /label= Asn, Ser
 XX
 XX
 PN US2002061562-A1.
 XX
 XX 23-MAY-2002.
 XX
 XX 09-AUG-2001; 2001US-00927602.
 XX
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX
 XX (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX
 XX Fukuda MN, Akama TO;
 PI
 XX WPI; 2002-507643/54.
 DR
 XX
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.

Example 5; Fig 2A-B; 69pp; English.

XX The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratinoplasty or keratotomy. The present sequence
 CC represents a consensus N-acetylglucosamine-6-sulfotransferase which is
 CC given in the exemplification of the present invention

XX Sequence 395 AA;

Query Match 85.7%; Score 1758.5; DB 5; Length 395;
 Best Local Similarity 87.1%; Pred. No. 1.6e-183;
 Matches 338; Conservative 10; Mismatches 39; Indels 1; Gaps 1;

QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
 DB 1 MWLPFRSSTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 59
 QY 61 FSQHPDVFYLMPEAHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDYMPPQSRNLSA 120
 DB 60 FSQHPDVFYLMPEAHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVDFDYMPPQSRNLSA 119
 QY 121 FFWNATSRALCSPACSAFPGRTISKQDVCKTCTROPFSIAREACSHYHVLKEVRF 180
 DB 120 LFQWANSRALCSPACSAFPGRTISKQDVCKTCTROPFSIAREACSHYHVLKEVRF 179
 QY 181 NLQVLYPLSDPALNLRIHVLPDRAVLSRREAGPILARDNGIVLGTNGKVEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIHVLPDRAVLSRREAGPILARDNGIVLGTNGKVEADPHL 239
 QY 241 RLIREVCRSHVRIAEAAATLKPFPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300
 DB 240 RVREVCVSHVRIAEAAATLKPFPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLPQL 299
 QY 301 EAMHNTHGSGCKPIEAFHTSSRNARNYSQAWRHALPTKILRVQEVGCAGALQLGYR 360
 DB 300 EAMHNTHGSGCKPIEAFHTSSRNARNYSQAWRHALPTKILRVQEVGCAGALQLGYR 359
 QY 361 PVYSADQOORDLTLDLVLPRGPDHFWAS 388
 DB 361 PVYSADQOORDLTLDLVLPRGPDHFWAS 387

Db 360 PVYSADQOORDLTLDLVLPRGPDHFWAS 387
 RESULT 5
 AAAY72640
 ID AAAY72640 standard; protein; 395 AA.
 XX
 XX AAAY72640;
 XX
 XX 02-MAY-2001 (first entry)
 DT
 XX
 XX Human glycosyl sulfotransferase-4beta (GST-4beta).
 DE
 XX
 XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX
 XX Homo sapiens.
 OS
 XX
 XX WO200106015-A1.
 PN
 XX
 XX 25-JAN-2001.
 PD
 XX
 XX 19-JUL-2000; 2000WO-US019741.
 XX
 XX 20-JUL-1999; 99US-0144694P.
 PR
 XX 13-JUN-2000; 2000US-00593828.
 PR
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Rosen SD, Lee JK, Hemmerich S;
 PI WPI; 2001-138471/14.
 XX N-PSDB; AAD02697, AAD02700.
 DR
 XX
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 PT
 XX Claim 3; Fig 4B; 128pp; English.
 PS
 XX
 XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 XX Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 4; Length 395;

Best Local Similarity 85.8%; Pred. No. 2.4e-180;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPFRSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
 DB 1 MWLPFRSSTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSSFLGQL 59

Db 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFVQL 59
Qy 61 FSOHPDVYLMPEPAHWVTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
Db 60 FNQHPDVYLMPEPAHWVTTLSQSSAATLHMAVRDLVRSVFLCMDVDFDAYLPWRRNLSD 119
Qy 121 FFWNATSRALCSPACSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179
Qy 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 240
Db 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 239
Qy 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRPDLAREPLAIRALYFTGLTLPQL 300
Db 240 RVREVCRSVRIAEATLKPPLRGYRLVRPDLAREPLAIRALYFTGLTLPQL 299
Qy 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOVCAGALQLLGYR 360
Db 300 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOVCAGALQLLGYR 359
Qy 361 PVYSADQORDLTLDLVLPRGDPHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

RESULT 6
ABB81554
ID ABB81554 standard; protein; 395 AA.
XX AC ABB81554;
XX DT 05-SEP-2002 (first entry)
XX DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
XX KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
XX KW ophthalmological; chromosome 16q22.
XX OS Homo sapiens.
XX PN US2002061562-A1.
XX FD 23-MAY-2002.
XX PF 09-AUG-2001; 2001US-00927602.
XX PR 11-AUG-2000; 2000US-00638211.
XX PR 11-AUG-2000; 2000US-0325773P.
XX PA (FUKU//) FUKUDA M N.
XX PA (AKAW//) AKAWA T O.
XX PI Fukuda MN, Akama TO;
XX DR WPI; 2002-507643/54.
XX DR N-PSDB; ABN89506.
XX PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX PT useful for treatment, monitoring and diagnosis of macular corneal
XX PT dystrophy.
XX PS Claim 13; Fig 1A-D; 69pp; English.
XX CC The present sequence represents human corneal N-acetylglucosamine-6-
XX CC sulfotransferase (I), which is able to catalyze sulfation of keratan
XX CC sulfate (KS). Also described is a method for monitoring the effect of
XX CC treatments for macular corneal dystrophy (MCD), and detecting
XX CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
XX CC ophthalmological activity. (I) can be used to treat or prevent macular
XX CC corneal dystrophy types I or II. (I) makes possible treatment of MCD

CC without requiring keratinoplasty or keratectomy
XX SQ Sequence 395 AA;
Query Match 84.3%; Score 1729.5; DB 5; Length 395;
Best Local Similarity 85.8%; Pred. No. 2.4e-180;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
Qy 1 MWLPRVSSKTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFVQL 60
Db 1 MWLPRVSSKTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVVLSSWRSGSSFVQL 59
Qy 61 FSOHPDVYLMPEPAHWVTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
Db 60 FNQHPDVYLMPEPAHWVTTLSQSSAATLHMAVRDLVRSVFLCMDVDFDAYLPWRRNLSD 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSAVCKPLCARQSFTLAREACRSYSHVVLKEVRF 179
Qy 181 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 240
Db 180 NLQVLYPLSDPALNLRIHVLRDPRAVLRSPAGGEARVHVVLSSWRSGSSFVQL 239
Qy 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRPDLAREPLAIRALYFTGLTLPQL 300
Db 240 RVREVCRSVRIAEATLKPPLRGYRLVRPDLAREPLAIRALYFTGLTLPQL 299
Qy 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOVCAGALQLLGYR 360
Db 300 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOQWRHALPFTKILRVOVCAGALQLLGYR 359
Qy 361 PVYSADQORDLTLDLVLPRGDPHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

RESULT 7
AAE15438
ID AAE15438 standard; protein; 395 AA.
XX AC AAE15438;
XX DT 12-MAR-2002 (first entry)
XX DE Human drug metabolising enzyme (DME)-5.
XX KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
XX KW inflammatory disorder; acquired immune deficiency syndrome; infection;
XX KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
XX KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
XX KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
XX KW gastrointestinal disorder; metabolic disorder; developmental disorder;
XX KW liver disorder; iritis; cystic fibrosis; Addison's disease; reinitis;
XX KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
XX KW DME-5.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..35 /label= Signal_peptide
XX FT Peptide 1..32 /label= Signal_peptide
XX FT Protein 33..395 /note= "Human mature DME-5 protein"
XX FT Protein 36..395 /note= "Human mature DME-5 protein"
XX WO200179468-A2.
XX 25-OCT-2001.

PF 12-APR-2001; 2001WO-US011869.
XX
PR 13-APR-2000; 2000US-0197590P.
PR 19-APR-2000; 2000US-0198403P.
PR 28-APR-2000; 2000US-0200185P.
PR 05-MAY-2000; 2000US-0202234P.
PR 11-MAY-2000; 2000US-0203509P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
PI Au-Young J;
XX
DR WPI; 2002-066363/09.
DR N-PSDB; AAD24670.
XX
XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
PT useful for diagnosing, treating, or preventing disorders associated with
PT aberrant expression of DME such as allergy, anemia, asthma, infertility.
XX
XX Claim 1a; Page 131-132; 143pp; English.
XX
XX The invention relates to human drug metabolising enzymes referred as DME
CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the
CC invention are useful for assessing toxicity of test compounds and in gene
CC therapy. Sequences of the invention are useful in the diagnosis,
CC prevention and treatment of autoimmune/inflammatory disorders such as
CC acquired immune deficiency syndrome (AIDS), adult respiratory distress
CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune
CC haemolytic anaemia, contact dermatitis, Crohn's disease,
CC glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease,
CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,
CC rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral,
CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell
CC proliferative disorders such as actinic keratosis, arteriosclerosis,
CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's
CC syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine
CC disorders such as disorders of the hypothalamus and pituitary resulting
CC from lesions such as primary brain tumours, adenomas, infarction
CC associated with pregnancy, aneurysms, vascular malformations; eye
CC disorders such as conjunctivitis, iritis, retinitis, glaucoma; pigmentosa
CC ; metabolic disorders such as Addison's disease, cystic fibrosis,
CC diabetis, goitre, glycogen storage diseases, hypercholesterolemia,
CC mannoidosis, obesity; gastrointestinal disorders such as dysphagia,
CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,
CC gastric carcinoma, anorexia, nausea, gastroenteritis,
CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
CC disorders. The present sequence is human DME-5 protein
XX
SQ Sequence 395 AA;

Query Match
Best Local Similarity 84.3%; Score 1729.5; DB 5; Length 395;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRFSSKTVTVLLAQTTCLLFTIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db 1 MWLPRVSTAVTALLAQ-TFLLFLVSRPGSPAGGEARVHVLVLSWRSGSFLGQL 59
Qy 61 FSDHPDVFYLMPEAHVHTTLLSGSAAATLHMAVRDLMSRIFLCMDVFDYMPQSRNLSA 120
Db 60 FNDHPDVFYLMPEAHVHTTLLSGSAAATLHMAVRDLVRSVFLCMDVFDYLAFLPWRNLS 119
Qy 121 FFWNATSRALCSPACSAFPRGTITSKQDVCKTLCITQPPSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSAEVCKPCARQSFLLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRAAGPILARDNGIVLGTNGKWEADPHL 240

Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRQETAKALARDNGIVLGTNGTWVEADPGL 239
Qy 241 RLIREVCRSHVRIAEAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 300
Db 240 RVREVCRRSHVRIAEAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 299
Qy 301 EAWIHNIHSGSIGKPIEAFHTSSRNARNVSAWRHALPFTKILRVQEVCAQALQLGYR 360
Db 300 EAWIHNIHSGSGGARREAFKTSRRNALNVSAWRHALPFAKIRRVQELCAGALQLGYR 359
Qy 361 PYSADQQRDLTDLVLPGRPDHFSWAS 388
Db 360 PVSSEDEQRNLADLDLVLPRGLNGFTWAS 387

RESULT 8
AD121086
ID AD121086 standard; protein; 395 AA.
XX
AC AD121086;
XX
DT 15-APR-2004 (first entry)
XX
DE Novel human protein #61.
XX
XX forensic; nutritional source; damaged tissue; diseased tissue;
KW myeloid cell disorder; lymphoid cell disorder;
KW bone cartilage tissue growth; tendon tissue growth;
KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;
KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.
XX
OS Homo sapiens.
XX
PN WO2003025148-A2.
XX
PD 27-MAR-2003.
XX
PF 19-SEP-2002; 2002WO-US029964.
XX
PR 19-SEP-2001; 2001US-0323739P.
PR 13-SEP-2002; 2002US-00323739.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;
PI Haley-Vicente D;
XX
XX WPI; 2003-354603/33.
XX N-PSDB; AD121802.
XX
XX New polynucleotides and secreted proteins, useful for treating myeloid or
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
PT tissue growth or regeneration, in wound healing, and in tissue repair and
PT replacement.
XX
XX Claim 20; SEQ ID NO 337; 156pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a
CC polypeptide with biological activity. The polynucleotides and
CC polypeptides are useful in diagnostics, forensics, gene mapping,
CC identification of mutations responsible for genetic disorders and other
CC traits, to assess biodiversity, as nutritional sources or supplements.
CC The polynucleotides may also be used as molecular weight markers,
CC chromosome markers or map related gene positions, or as an antigen to
CC raise anti-DNA antibodies or elicit immune response. The polypeptides are
CC useful for raising antibodies, as markers for tissues in which the
CC corresponding polypeptide is expressed, for re-engineering damaged or
CC diseased tissues, for treating myeloid or lymphoid cell disorders, in
CC bone cartilage, tendon, ligament and/or nerve tissue growth or
CC regeneration, in wound healing, in tissue repair and replacement, in
CC healing of burns, incisions and ulcers, and in treating cancer. The
CC present sequence represents the amino acid sequence of a novel human

CC protein.
 XX
 SQ Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 7; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37;

QY 1 MWLPRFSSKTVTVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
 DB 1 MWLPRVSSSTAVTALLAQ-TFLLFLVSRPSPAGGEARVHVLVLSWSSGSSFGQL 59

QY 61 FSQHPDVFYLMPEAHVHTTLLSQSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNQHPDVFYLMPEAHVHTTLLSQSAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNLS 119

QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKEVRFF 179

QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVADPGL 239

QY 241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 300
 DB 240 RVREVCRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 299

QY 301 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILURVOECAGALQLLYR 360
 DB 300 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILURVOECAGALQLLYR 359

QY 361 PVYSADOQRDLTLDLVLPRGPDHFSWAS 388
 DB 360 PVYSEDEQRNALDLDLVLPRGLNGFTWAS 387

RESULT 9
 ADL61235
 ID ADL61235 standard; protein; 395 AA.
 XX
 AC ADL61235;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
 KW predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
 KW vasotrophic; vulnary; pharmacogenomic; drug sensitivity; breast cancer;
 KW hypervascular disease; angiogenesis; wound healing scar; human;
 KW biomarker; carbohydrate sulphotransferase 6; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN WO2004020583-A2.
 XX
 PD 11-MAR-2004.
 XX
 PF 26-AUG-2003; 2003WO-US026491.
 XX
 PR 27-AUG-2002; 2002US-0406385P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Huang F, Han X, Reeves KA, Amier L, Fairchild CR, Lee FY;
 PI Shaw P;
 XX
 DR WPI; 2004-239171/22.
 DR N-PSDB; ADL61098.
 XX

New predictor sets with a plurality of polynucleotides and/or
 polypeptides whose expression pattern predicts cell response to a
 compound that modulates protein tyrosine kinase activity, useful in

PT treating breast cancer.
 XX
 PS Claim 9; SEQ ID NO 159; 649pp; English.
 XX

The invention relates to a novel predictor set comprising a plurality of
 CC polynucleotides and/or polypeptides whose expression pattern is
 CC predictive of the response of cells to treatment with a compound that
 CC modulates protein tyrosine kinase activity or members of the protein
 CC tyrosine kinase pathway. The molecules of the invention demonstrate
 CC cytostatic, antiangiogenic, vasotrophic and vulnary activities and may
 CC be useful in the field of pharmacogenomics, in particular for determining
 CC drug sensitivity and in treating breast cancer, hypervascular diseases,
 CC angiogenesis and scars in wound healing. The current sequence is that of
 CC a human protein tyrosine kinase biomarker protein of the invention.
 XX
 SQ Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 8; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.4e-180; Indels 1; Gaps 1;
 Matches 333; Conservative 17; Mismatches 37;

QY 1 MWLPRFSSKTVTVLLAQTCLLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
 DB 1 MWLPRVSSSTAVTALLAQ-TFLLFLVSRPSPAGGEARVHVLVLSWSSGSSFGQL 59

QY 61 FSQHPDVFYLMPEAHVHTTLLSQSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNQHPDVFYLMPEAHVHTTLLSQSAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRNLS 119

QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFQWAVSRALCSPACSAFPRGAISSBAVCKPLCARQSFTLAREACRSYSHVVLKEVRFF 179

QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWVADPGL 239

QY 241 RLIREVCRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 300
 DB 240 RVREVCRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 299

QY 301 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILURVOECAGALQLLYR 360
 DB 300 EAWIHNTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILURVOECAGALQLLYR 359

QY 361 PVYSADOQRDLTLDLVLPRGPDHFSWAS 388
 DB 360 PVYSEDEQRNALDLDLVLPRGLNGFTWAS 387

RESULT 10
 AAY72638
 ID AAY72638 standard; protein; 395 AA.
 XX
 AC AAY72638;
 XX
 DT 02-MAY-2001 (first entry)
 XX
 DE Mouse glycosyl sulfotransferase-4 (GST-4).
 XX
 KW Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;
 KW selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailtis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 8E1.
 XX
 OS Mus musculus.
 XX

PN WO200106015-A1.
 XX 25-JAN-2001.
 XX 19-JUL-2000; 2000WO-US019741.
 XX 20-JUN-1999; 99US-0144694P.
 PR 13-JUN-2000; 2000US-00593828.
 XX (REGC) UNIV CALIFORNIA.
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI; 2001-138471/14.
 DR N-PSDB; AAD02696.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX Claim 3; Fig 2; 128pp; English.
 XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
 CC gene is found on chromosome 8E1. GST is a type 2 membrane protein useful
 CC for inhibiting a binding event between a selectin and a selectin ligand,
 CC which comprises contacting the selectin with a non-sulphated selectin
 CC ligand, GST and a small molecular agent that inhibits the sulphation
 CC activity of GST. GST is also useful in inhibiting a selectin mediated
 CC binding event. GST is useful in gene therapy to treat disorders such as
 CC acute or chronic inflammation, systemic lupus erythematosus (SLE),
 CC rheumatoid arthritis, polyarteritis nodosa, polymyositis,
 CC dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
 CC myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
 CC disease, adrenalitis, hypoparathyroidism, pernicious anaemia,
 CC demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
 CC myocarditis, regional enteritis, adult respiratory distress syndrome,
 CC infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
 CC asthma, hypersensitivity, rheumatic fever and tissue rejection during
 CC transplantation
 XX Sequence 395 AA;
 SQ

Query Match 74.9%; Score 1536.5; DB 4; Length 395;
 Best Local Similarity 76.0%; Pred. No. 3.6e-159;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRFSSTVTVLLAQTCLLFIISRPGSPAGGEDRVHVLVLSWRSSSFFVQL 60
 Db 1 MRLPRFSSTVTVLLMVQTG-ILVFLVSRQVPSPPAGLGRVHVLVLSWRSSSFFVQL 59

Qy 61 FSOHPDVYLMPEAWHVTTLTSSQSAATLHMAVRDLMSIFLDCMDVDFDYMPSQSNLSA 120
 Db 60 FSOHPDVYLMPEAWHVTTLTSSQSAALHMAVRDLMSIFLDCMDVDFDYMPSQSNISD 119

Qy 121 FFWNATSRALCSPACSAFPRGTISKDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
 Db 120 LFWNATSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 179

Qy 181 NLQVLYPLLSDPALNLRIVHLRDPRAVLRSREAGPILARDNGIVLGTGKWEADPHL 240
 Db 180 NLQVLYPLLSDPALNLRIVHLRDPRAVLRSREAGPILARDNGIVLGTGKWEADPHL 239

Qy 241 RLIREVCRSHVRAEATLPPPPPLRCRYRLRVFEDLAREPLAEIRALYFTGLTLPQL 300
 Db 240 RVNEVCRSHVRAEALHPPPPPLQDRYRLRVFEDLAREPLTVIRELYFTGLTLPQL 299

Qy 301 EAWIHNITHSGIGKPIEAFHTSSRNARNYSQAWRHALPFTKILRVOEVACAGALQLGYR 360
 Db 300 QTIWNIHITHSGPGARAEAKFTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLGYR 359

Qy 361 PVSADQORDLTDLVLPGRGDPHSWAS 388
 Db 360 SVHSELEQDLSLDLLPRGMDSPKAS 387

RESULT 11
 AAU11275
 ID AAU11275 standard; protein; 395 AA.
 XX AAU11275;
 XX 12-MAR-2002 (first entry)
 XX Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.
 XX Mouse; beta1,3GnT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
 KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
 KW anti-inflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
 XX OS Mus musculus.
 XX WO200185177-A1.
 XX 15-NOV-2001.
 XX 10-MAY-2001; 2001WO-US015452.
 XX 11-MAY-2000; 2000US-00569320.
 XX (BURN-) BURNHAM INST.
 XX Fukuda M, Yeh J, Hiraoka N;
 WPI; 2002-075226/10.
 N-PSDB; AAS16948.
 New enzyme, useful for modifying acceptor molecule, comprises an isolated
 L-selectin sulfotransferase-2 that directs expression of L-selectin
 ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
 GlcNAc 6-sulfotransferase.
 Claim 28; Fig 10; 98pp; English.
 The present invention provides a method of modifying an acceptor molecule
 by contacting the acceptor with an isolated beta1,3-N-
 acetylglucosaminyltransferase (beta1,3GnT) or an active fragment, where
 beta1,3GnT directs expression of a MECA-79 antigen. The invention also
 provides a method of treating or preventing an L-selectin-mediated
 condition by reducing the expression or activity of a beta1,3GnT that
 directs expression of a MECA-79 antigen. This can be done by
 administering to the subject an oligosaccharide L-selectin antagonist
 that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 administering antibody material that specifically binds beta1,3GnT,
 and/or a beta1,3GnT antisense nucleic acid molecule. L-selectin
 sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
 Alternatively, the expression or activity of LSST-2 or its active
 fragment can be reduced in combination with reducing the expression or
 activity of beta1,3GnT. The method is useful for treating L-selectin
 mediated conditions such as Crohn's disease and ulcerative colitis,
 inflammatory disorders of the skin such as allergic contact dermatitis,
 psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 hypersensitivity reactions, diabetes and hyperplastic thymus. This
 sequence represents mouse I-GlcNAc6ST

SQ Sequence 395 AA;
 Query Match 74.9%; Score 1536.5; DB 5; Length 395;
 Best Local Similarity 76.0%; Pred. No. 3.6e-159;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRFSSTVTVLLAQTCLLFIISRPGSPAGGEDRVHVLVLSWRSSSFFVQL 60
 Db 1 MRLPRFSSTVTVLLMVQTG-ILVFLVSRQVPSPPAGLGRVHVLVLSWRSSSFFVQL 59

Qy 61 FSQHPDVLYLMEPAHWVWTLTSSQSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLNSA 120
 Db 60 FSQHPDVLYLMEPAHWVWTLTSSQSAAPALHMAVRDLMSIFLCMDVDFDAYMPOSRLNSA 119
 Qy 121 FFWATSRALCSPACAPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
 Db 120 LFWAVSRALCSPVCEAFAGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 179
 Qy 181 NLQVLYPLSDPALNLRIVHLVRDPAVLSREAAGPILARDNGIVLGTNGKWEADPHL 240
 Db 180 NLQVLYPLSDPALNLRIVHLVRDPAVLSREAAGPILARDNGIVLGTNGKWEADPHL 239
 Qy 241 RLIREVCRSHVRIAEAAATLKPPPLGRYRLVRPDLAREPLAEIRALYFTGLTLPQL 300
 Db 240 RVNEVCRSHVRIAEAAATLKPPPLGRYRLVRPDLAREPLAEIRALYFTGLTLPQL 299
 Qy 301 EAWHNTTHSGGIGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOEVCAGALQLLGYR 360
 Db 300 QTWIHNTHSGGPGARREAFKTSRDALSVSQAWRHLPFTKILRVOEVCAGALQLLGYR 359
 Qy 361 PVYSADOORDLTLDLVLPGRGDHFSWAS 388
 Db 360 SVHSELRDLSLDDLLPRGMDSPFKWAS 387

RESULT 12
 ABB81557
 ID ABB81557 standard; protein; 418 AA.
 AC ABB81557;
 DT 05-SEP-2002 (first entry)
 DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 OS Mus musculus.
 PN US2002061562-A1.
 PD 23-MAY-2002.
 PF 09-AUG-2001; 2001US-00927602.
 PR 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX (FUKU/) FUKUDA M N.
 XX (AKAM/) AKAMA T O.
 XX Fukuda MN, Akama TO;
 XX WPI; 2002-507643/54.
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX Example 5; Page 24-25; 69pp; English.

CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratotomy. The present sequence
 CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,

CC which is given in comparison with (I) in the exemplification of the
 CC present invention
 XX
 SQ Sequence 418 AA;
 Query Match 74.3%; Score 1523; DB 5; Length 418;
 Best Local Similarity 75.8%; Pred. No. 1.2e-157;
 Matches 294; Conservative 30; Mismatches 62; Indels 2; Gaps 2;
 Qy 1 MWLPRSSKTVTVLLLAQTCLLFIISRRPGSPAGGEDRVHVLVLSRSGSFFLQQL 60
 Db 25 MWLPRSSKTVTVLLLAQTCLLFIISRRPGSPAGGEDRVHVLVLSRSGSFFLQQL 83
 Qy 61 FSQHPDVLYLMEPAHWVWTLTSSQSAATLHMAVRDLMSIFLCMDVDFDAYMPOSRLNSA 120
 Db 84 FSQHPDVLYLMEPAHWVWTLTSSQSAAPALHMAVRDLMSIFLCMDVDFDAYMPOSRLNSA 143
 Qy 121 FFWATSRALCSPACAPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
 Db 144 LFWAVSRALCSPVCEAFAGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 203
 Qy 181 NLQVLYPLSDPALNLRIVHLVRDPAVLSREAAGPILARDNGIVLGTNGKWEADPHL 240
 Db 204 NLQVLYPLSDPALNLRIVHLVRDPAVLSREAAGPILARDNGIVLGTNGKWEADPHL 263
 Qy 241 RLIREVCRSHVRIAEAAATLKPPPLGRYRLVRPDLAREPLAEIRALYFTGLTLPQL 300
 Db 264 RVNEVCRSHVRIAEAAATLKPPPLGRYRLVRPDLAREPLAEIRALYFTGLTLPQL 322
 Qy 301 EAWHNTTHSGGIGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOEVCAGALQLLGYR 360
 Db 323 QTWIHNTHSGGPGARREAFKTSRDALSVSQAWRHLPFTKILRVOEVCAGALQLLGYR 382
 Qy 361 PVYSADOORDLTLDLVLPGRGDHFSWAS 388
 Db 383 SVHSELRDLSLDDLLPRGMDSPFKWAS 410

RESULT 13
 ADJ70405
 ID ADJ70405 standard; protein; 394 AA.
 AC ADJ70405;
 DT 06-MAY-2004 (first entry)
 DE Human heat mitochondrial protein as a therapeutic target SeqID2211.
 KW mitochondrial; human; screening assay; diabetes mellitus;
 KW Huntington's disease; osteoarthritis;
 KW Leber's hereditary optic neuropathy; LHON;
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
 KW osteopathic; ophthalmological; cytostatic.
 OS Homo sapiens.
 XX WO2003087768-A2.
 XX 23-OCT-2003.
 XX 04-APR-2003; 2003WO-US010870.
 XX 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-0389987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX (MITO-) MITOKOR.
 XX (BUCK-) BUCK INST AGE RES.
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DB;

XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 PS Claim 1; SEQ ID NO 2211; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, nootropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cyostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 SQ Sequence 394 AA;

Query Match 73.8%; Score 1513.5; DB 7; Length 394;
 Best Local Similarity 92.6%; Pred. No. 1.2e-156;
 Matches 289; Conservative 7; Mismatches 15; Indels 1; Gaps 1;
 QY 1 MMLPRSSKTVLLIAQTCLLFIISRGSPSPAGGEDRVHVLVLSWSRSGSFLGQL 60
 Db 1 MMLPRVSSSTAVTALLAQ-TFLLLFLVSRGSPSPAGGEARVHVLVLSWSRSGSFLGQ 59
 QY 61 FSOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVARDLMRSIFLCMDMDFDAYMPSQRNLSA 120
 Db 60 FNQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVARDLRSVFLCMDVDFDAYLWRRNLS 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTLCARQPFSLAREACKRSYSHVVLKEVRFF 180
 Db 120 LFQWAVSRALCSPSPACSAFPRGAISKQDVCKTLCARQPFSLAREACKRSYSHVVLKEVRFF 179
 QY 181 NLQVLYPLLSDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKVEADPHL 240
 Db 180 NLQVLYPLLSDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTNGKVEADPHL 239
 QY 241 RLIREVCRSHVRIAEEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 300
 Db 240 RLIREVCRSHVRIAEEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 299
 QY 301 EAWIHNITHGSG 312
 Db 300 EAWIHNITHGSG 311

RESULT 14
 AAM93309 standard; protein; 386 AA.
 ID AAM93309
 XX
 AC AAM93309;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 2817.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 FN EPI130094-A2.
 XX

PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayaishi K, Ishii S, Kawai Y;
 PI Wakanatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 WPI; 2001-524255/58.
 DR N-PSDB; AAK94229.
 XX
 XX 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PS Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX
 SQ Sequence 386 AA;
 Query Match 49.4%; Score 1013; DB 4; Length 386;
 Best Local Similarity 52.5%; Pred. No. 9.9e-102;
 Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;
 QY 14 LLLAQTCTLLFIISRP-----GFSSPAGGEDRVHVLVLSWSRSGSFLGQ 59
 Db 1 MLLPKKMLLLFLVSMQAILALFFHMYSHNISLSMKAPERMHVLVLSWSRSGSFLGQ 60
 QY 60 LFSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVARDLMRSIFLCMDMDFDAYM-POSRL 118
 Db 61 LFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVARDLMRSIFLCMDMDFDAYM-POSRL 120
 QY 119 SAFFNATSRALCSPSPACSAFPRGTISKQDVCKTLCARQPFSLAREACKRSYSHVVLKEVR 178
 Db 121 SFLFQWENSRALCSPACDIIIPQDEIIIPRAHCELLCSQPFVVEKACRSYSHVVLKEVR 180
 QY 179 FPNLQVLYPLLSDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWKEAD 237
 Db 181 FPNLQVLYPLLSDPALNLRIHVLVRDPRAVLRSREAAAGPILARDNGIVLGTN-GKWKEAD 240
 QY 238 PHLLIREVCRSHVRIAEEAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLTL 296
 Db 241 QPYVNMQVTCQSOLEIYK--TIQSLPKALOERLLVRYEDLAPAPVQAQSRMTEFFVGLF 298
 QY 297 TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSQAWRHALPFTTKILRVQEVYACAGALQL 356
 Db 299 LPHLQVWVNIITGKMGD--HAFHTNARDALNVSQAWRWSLPEYKVRSLQKACGDAMNL 356
 QY 357 LGYRVPYSDAQDQDLTLDLVLPRGPDHFSWASPD 390
 Db 357 LGYRHRVSRSEQORNLLDLL-----STWTVP 383
 RESULT 15
 ADL30784 standard; protein; 386 AA.
 ID ADL30784
 XX

AC ADL30784;
XX 20-MAY-2004 (first entry)
XX Human protein encoded by a full length cDNA clone seqID 2817.
XX human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method.
XX Homo sapiens.
XX EPI396543-A2.
XX 10-MAR-2004.
XX 07-JUL-2000; 2003EP-00025638.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
FT length human cDNAs.
XX Example 1; SEQ ID NO 2817; 1340pp; English.

XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
XX length human protein of the invention.
SQ Sequence 386 AA;

Query Match 49.4%; Score 1013; DB 8; Length 386;
Best Local Similarity 52.5%; Pred. No. 9.9e-102;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;
Qy 14 LLLAQTTCLLLPIISRP-----GPSPPAGGEDRVHVLSSWRSGSSFLGQ 59
Db 1 MLLPKMKMLLLFLVSQMAILALFFHMYSHNTSSLSMKAPQPMHVLSSWRSGSFVQG 60
Qy 60 LFSQHPDVFYLMPEAWHVMYTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYM-PQSRNL 118
Db 61 LFGQHPDVFYLMPEAWHVMYTTKSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEGPRRQ 120
Qy 119 SAFFNWTASRALCSPACAPFRGTISKQVCKTLCTQPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWNSRALCSPACADIIIPQDEIIIPRAHCLICSQPFVEVKEKACRSYSHVVLKEVR 180
Qy 179 FPNQLVYPLLSDDPALNIRIVHLVRDPRAVLRREAAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 FPNQLVYPLLSDDPALNIRIVHLVRDPRAVLRREAAGPILARDNGIVLGTN-GKWVEAD 240
Qy 238 PHRLRIREVCSHVRVIAEATLKP-PPLRGYRLVRFEDLAREPLAEIRALYAFGLTL 296
Db 241 QPYVMQVICSQLEIYK--TIQSLPKALQRYLLVRYEDLARAPVAQTSRMIEFVGLEF 298
Qy 297 TPQLEAWIHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQL 356

Db 299 LPHLQTVVHNITRGKMGD--HAFHTNARDALNVSOAWRHSLPYEKVSRQLKACGDAMNL 356
Qy 357 LGYRPVYSADQQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRHYRSEQEQRNLLDLL-----STWTVPE 383
Search completed: June 23, 2005, 08:43:23
Job time : 43.1481 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 42.7985 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPRSSKTVTLVLLAQTT.....LTLDLVLPRGDFHFWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2051	100.0	390	2	Q7LCN3 homo sapien
2	2051	100.0	411	2	Q9GZS9
3	1729.5	84.3	395	2	Q9GZK3
4	1536.5	74.9	395	2	Q9QUR4
5	1119.5	54.6	392	2	Q6GN39
6	1013	49.4	386	2	Q9NCG5
7	1008	49.1	386	2	Q9Y5R3
8	999	48.7	370	2	Q9IV46
9	965.5	47.1	388	2	Q9RI11
10	958.5	46.7	388	2	Q9WUES
11	665	32.4	483	2	Q794G9
12	665	32.4	530	2	Q88276
13	662	32.3	530	2	Q80WV3
14	659.5	32.2	483	2	Q9UEU5
15	659.5	32.2	530	2	Q9Y4C5
16	617.5	30.1	484	2	Q9EP78
17	614.5	30.0	484	2	Q9N8B0
18	610	29.7	486	2	Q75667
19	609	29.7	486	2	Q9NS84
20	603	29.4	485	2	Q6XQG8
21	579.5	28.3	479	2	Q7LGC8
22	577.5	28.2	479	2	Q75099
23	576	28.1	472	2	Q88199
24	560	27.3	474	2	Q9QZL2
25	552	26.9	420	2	Q6DBY9
26	540.5	26.4	458	1	C8ST CHICK
27	520	25.4	411	2	Q9EQC0
28	513.5	25.0	411	2	Q43916
29	472.5	23.0	441	2	Q93403
30	403.5	19.7	257	2	Q79415
31	353.5	17.2	225	2	Q6RY62

Q677T4 lymphocveti
Q9VMC3 drosophila
Q7QIF9 anopheles g
Q95J48 tryctolagus
Q9VMC4 drosophila
Q9D0K5 mus musculus
Q8MZD1 drosophila
Q9JJE6 streptomyce
Q81ZU8 homo sapien
Q7V4N3 prochloroco
Q92VG4 rhizobium m
Q68RS5 prochloron
Q8PRA0 xanthomonas
Q6ZET4 synechocyst

ALIGNMENTS

RESULT 1

Q7LCN3 O7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Name-I-GlcNAc-6-ST;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA MEDLINE=9443499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RX Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
sulfotransferase that is highly restricted to intestinal tissue.";
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1; -
DR EMBL; AF176838; AAD56000.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 100.0%; Score 2051; DB 2; Length 390;

Best Local Similarity 100.0%; Pred. No. 3.3e-175;

Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MWLPRSSKTVTLVLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWSGSSFLGQL 60
Db 1 MWLPRSSKTVTLVLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWSGSSFLGQL 60
Qy 61 FSOHPDVFYLMPEPAHWVTTLSQGSAAATLHMAVDLMRSIFCLDMDVFDAYMPQSRNL 120
Db 61 FSOHPDVFYLMPEPAHWVTTLSQGSAAATLHMAVDLMRSIFCLDMDVFDAYMPQSRNL 120
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDCKTCTCTROPFSLAREACRSYSHVVLKEVRF 180
Db 121 FFWNATSRALCSPACSAFPRGTISKQDCKTCTCTROPFSLAREACRSYSHVVLKEVRF 180
Qy 181 NLQVLYLLSDPALNLRIVHLVRDPAVLRSLREAGPILARDNGILVTGNGKWEADPHL 240
Db 181 NLQVLYLLSDPALNLRIVHLVRDPAVLRSLREAGPILARDNGILVTGNGKWEADPHL 240

```
QY 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 300
Db 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 300
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
Db 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
QY 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390

RESULT 2
Q9GZS9 PRELIMINARY; PRT; 411 AA.
ID Q9GZS9 Q9UB3;
AC Q9GZS9; Q9UB3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAC28023.1; -.
DR EMBL; AF219991; AAC26326.1; -.
DR Genbank; HGNC:1973; CHST5.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 100.0%; Score 2051; DB 2; Length 411;
Best Local Similarity 100.0%; Pred No. 3.6e-175;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPRFSKTVTVLLAQTTCLLLFIISRPSPSPAGGEDRVHVLSSWRSGSFLGQL 60
Db 22 MWLPRFSKTVTVLLAQTTCLLLFIISRPSPSPAGGEDRVHVLSSWRSGSFLGQL 81
QY 61 FSQHPDVFYLMPEAWHWTTLSQSAATLHMVARDLMRSIFCLDMVDVDFAYMPQSRNLSA 120
Db 82 FSQHPDVFYLMPEAWHWTTLSQSAATLHMVARDLMRSIFCLDMVDVDFAYMPQSRNLSA 141
QY 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTQPPFLSAREACRSYSHVVLKEVRFF 180
Db 142 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTQPPFLSAREACRSYSHVVLKEVRFF 201
QY 181 NLQVYLLSDPALNLRIVHLVRDPRAVLSREAAGPILARDNGIVLGTNGKWEADPHL 240
Db 202 NLQVYLLSDPALNLRIVHLVRDPRAVLSREAAGPILARDNGIVLGTNGKWEADPHL 261
QY 241 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 300
Db 262 RLIREVCRSHVIAEATLPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLTTPQL 321
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 360
```

```
Db 322 EAWIHNTGSGIGKPIEAFHTSSRNARNVSQARHALPFTKILRVOECAGALQLLYR 381
QY 361 PVSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 382 PVSADQORDLTLDLVLPRGPDHFSWASPD 411

RESULT 3
Q9GZX3 PRELIMINARY; PRT; 395 AA.
ID Q9GZX3
AC Q9GZX3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase GSI-4beta (Corneal N-acetylglucosamine-6-O-sulfotransferase) (Carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6).
GN Name=GST4beta; Synonyms=CHST6;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family.";
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Guichwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywicki M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAG48244.1; -.
DR EMBL; AF219991; AAG26327.1; -.
```

EMBL; BC074883; AAH74883.1; -.
 DR EMBL; BC074834; AAH74834.1; -.
 DR EMBL; AF219990; AAG26325.1; -.
 DR Genew; HGNC.6938; CHST6.
 DR GO; GO:0005794; C:Golgi apparatus; TAS.
 DR GO; GO:0001517; F:N-acetylglucosamine 6-O-sulfotransferase ac. .; TAS.
 DR GO; GO:0006044; F:N-acetylglucosamine metabolism; IC.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;
 Query Match 84.3%; Score 1729.5; DB 2; Length 395;
 Best Local Similarity 85.8%; Pred. No. 2.2e-146;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;
 QY 1 MWLPRSSKTVTVLLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 DB 1 MWLPRVSSSTAVTALLAQ-TFLLLFLVSRPGPSPAGGEARVHVLVLSWRSGSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMVARDLVRSVFLCDMDVDFDAYLPWRRNLSD 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFWAVSRLCSPSPACSAFPRGAISSSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFF 179
 QY 181 NLQVLYPLLSDPALNLRIHVLVDRPRAVLRSREAGPIILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLLSDPALNLRIHVLVDRPRAVLRSREQTAKALARDNGIVLGTNGTWYVADPGL 239
 QY 241 RLIREVCRSHVRIAEATLKPPLGRGYRLVRPDLAREPLAEIRALYAFGLTGLTPQL 300
 DB 240 RVNVECRSHVRIAEATLKPPLGRGYRLVRPDLAREPLAEIRALYAFGLTGLTPQL 299
 QY 301 EAMTHNTHSGGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCCAGALQLLYR 360
 DB 300 EAMTHNTHSGGPGCARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLYR 359
 QY 361 PVYSADOQDRLTLDLVLPRGDPHFSWAS 388
 DB 360 PVYSEQRNLALDLVLPRLNGLGTWAS 387
 RESULT 4
 Q9QUP4 ID Q9QUP4 PRELIMINARY; PRT; 395 AA.
 AC Q9QUP4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase.
 GN Name=Chst5; Synonyms=I-GlcNAC-6-ST;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=C57Bl/6; TISSUE=Intestine;
 RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
 RX Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
 RT sulfotransferase that is highly restricted to intestinal tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57Bl/6; TISSUE=Intestine;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF176841; AAD56003.1; -.
 DR EMBL; AF176840; AAD56002.1; -.

MGD; MGI:1931825; Chst5.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Transferase.
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;
 Query Match 74.9%; Score 1536.5; DB 2; Length 395;
 Best Local Similarity 76.0%; Pred. No. 4.5e-129;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
 QY 1 MWLPRSSKTVTVLLLAQTCLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 DB 1 MRLPRFSSTVMLSULMVQTG-ILVFLVSRQVPSPPAGLGRVHVLVLSWRSGSFLGQL 59
 QY 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMVARDLMRSIFLCMDVDFDAYMPOSRLSA 120
 DB 60 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMVARDLVRSVFLCDMDVDFDAYLPWRRNLSD 119
 QY 121 FFWNATSRALCSPSPACSAFPRGTISKQDVCKTCTCTROPFSLAREACRSYSHVVLKEVRFF 180
 DB 120 LFWAVSRLCSPSPVCEAFARGNISSEVCKPLCATRPFGLAQEACRSYSHVVLKEVRFF 179
 QY 181 NLQVLYPLLSDPALNLRIHVLVDRPRAVLRSREAGPIILARDNGIVLGTNGKWEADPHL 240
 DB 180 NLQVLYPLLSDPALNLRIHVLVDRPRAVLRSREQTAKALARDNGIVLGTNGTWYVADPRL 239
 QY 241 RLIREVCRSHVRIAEATLKPPLGRGYRLVRPDLAREPLAEIRALYAFGLTGLTPQL 300
 DB 240 RVNVECRSHVRIAEATLKPPLGRGYRLVRPDLAREPLAEIRALYAFGLTGLTPQL 299
 QY 301 EAMTHNTHSGGIGKPIEAFHTSSRNARNVSQAWRHALPFTKILRVQEVCCAGALQLLYR 360
 DB 300 QTWIHNTHSGGPGCARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCAGALQLLYR 359
 QY 361 PVYSADOQDRLTLDLVLPRGDPHFSWAS 388
 DB 360 SVHSEQRDLSDLLPLPRGMDSPFKWAS 387
 RESULT 5
 Q6GN39 ID Q6GN39 PRELIMINARY; PRT; 392 AA.
 AC Q6GN39;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MGC83048 protein.
 GN Name=MGC83048;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RX MEDLINE=2338257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC073681; AAH73681.1; -;
GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
SQ SEQUENCE 392 AA; 45969 MW; 7712810F8B4704D7 CRC64;
Query Match 54.6%; Score 1119.5; DB 2; Length 392;
Best Local Similarity 55.6%; Pred. No. 1.1e-91;
Matches 217; Conservative 60; Mismatches 102; Indels 11; Gaps 5;
QY 3 LPRFSKTVTV--LLAQTTCLLFIISR----PGSPAGGEDRVHVLVLSWRSRSGSS 56
DB 1 MARFRPNVTVAGFLLQTIFLL-IYRHTVLPDTETK--EKVHLLLSWRSRSGSS 56
QY 57 LQGLFSQHPDVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIFLCMDVFDAYMPQSR 116
DB 57 IGOIFSQHPDVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIFLCMDVFDAYMPQSR 116
QY 117 NLSAFFNWATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKE 176
DB 117 NVSELFQWATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKE 176
QY 177 VRFNQLVYPLSDPALNLRIVHVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 236
DB 177 VRFNQLVYPLSDPALNLRIVHVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 236
QY 237 DPHLRILREVCVRSHVRIAEAAATLKP-PFPLRGYRLVRFEDLAREPLAEIRALYFTGLTL 296
DB 237 DPHLRILREVCVRSHVRIAEAAATLKP-PFPLRGYRLVRFEDLAREPLAEIRALYFTGLTL 296
QY 236 DIRYDVMREICQSHQAQMYETAMDKPSFLKGRYMLVRYEDVDRDPLREINQMTEFANLKL 295
QY 297 TPQLEAWIHNTGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKILRVOEVCAGALQL 356
DB 296 TAKLQWFFNITGVGPGTKKEEFTTSRNARNVSOAWRHALPFTKILRVOEVCAGALQL 355
QY 357 LGYRPVYSDQQRDLTLDLVRGPDHFSW 386
DB 356 LGYQFIDSEKEXKXMSDFLPRKRYQFSW 385
RESULT 6
Q8NCG5 PRELIMINARY; PRT; 386 AA.
AC Q8NCG5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ90265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,

RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074746; BAC11177.1; -;
DR Genew; HGNC:1972; CHST4.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase_1; 1.
RW Transferase.
SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;
Query Match 49.4%; Score 1013; DB 2; Length 386;
Best Local Similarity 52.5%; Pred. No. 3.7e-82;
Matches 207; Conservative 55; Mismatches 104; Indels 28; Gaps 7;
QY 14 LLLAQTTCCLLFIISR-----GPSPAGGEDRVHVLVLSWRSRSGSSFLGQ 59
DB 1 MLLPKKWKLLFLVVSQMAILALFFHMYSHNSSLMSKQAQPERMHVLSWRSRSGSSFLGQ 60
QY 60 LFSQHPDVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIFLCMDVFDAYM-POSRL 118
DB 61 LFSQHPDVFYLMPEAWHVWTTLSQGSAAATHMAVRDLMSIFLCMDVFDAYM-POSRL 120
QY 119 SAFFNWATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
DB 121 SFLFQWATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 180
QY 179 FNFNQLVYPLSDPALNLRIVHVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 237
DB 181 FNFNQLVYPLSDPALNLRIVHVRDPAVLRSAAGPILARDNGIVLGTN-GKWEAD 240
QY 238 PHLRLIREVCVRSHVRIAEAAATLKP-PFPLRGYRLVRFEDLAREPLAEIRALYFTGLTL 296
DB 241 QPYVNVQVICSQSLYK--TIQSLPKALQRYLLVRYEDLARAQVPAQTSRMTEFVGLF 298
QY 297 TPQLEAWIHNTGSGIGKPIEAFTSSRNARNVSOAWRHALPFTKILRVOEVCAGALQL 356
DB 299 LPHLQWFFNITGVGPGTKKEEFTTSRNARNVSOAWRHALPFTKILRVOEVCAGALQL 356
QY 357 LGYRPVYSDQQRDLTLDLVRGPDHFSW 390
DB 357 LGYRHRVSRSEQRNLLDLL-----STWTYPE 383
RESULT 7
Q9Y5R3 PRELIMINARY; PRT; 386 AA.
AC Q9Y5R3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN Name=GST3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;

```

RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family";
RL Glycobiology 11:75-87(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellices L.G.,
RA Rabuka D., Hindsgaul O., March J.D., Lowe J.B., Fakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT Core1 extension beta 1,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
DR ENBL; AF131235; AAD33015.1; -
DR ENBL; AF280088; AAG48246.1; -
DR ENBL; AF49783; AAK48417.1; -
DR GO; GO:0008146; F:sulfotransferase activity; TAS.
DR GO; GO:0007155; P:cell adhesion; TAS.
DR GO; GO:0006928; P:cell motility; TAS.
DR GO; GO:0007267; P:cell-cell signaling; TAS.
DR GO; GO:0006955; P:immune response; TAS.
DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Selectin; Transferase.
SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 49.1%; Score 1008; DB 2; Length 386;
Best Local Similarity 52.3%; Pred. No. 1e-81;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTTCLLLFIISRP-----GPSPPAGGEDRVVILVLSWRSGSSFLGQ 59
Db 1 MLLPKMKLLFLVLSQMAILALFFHMYSHNITSLSMKAQPERMHLVLSWRSGSSFVGQ 60
Qy 60 LFSQHPDVFYLMPEAWHWTTLSGSAATLHMVDRDLMSIFLDCMDVDFDAYM-PQGRNL 118
Db 61 LFGQHPDVFYLMPEAWHWMTFKSTAWMLHMVDRDLIRAVFLDCMSVDFDAYMFGPPRQ 120
Qy 119 SAFENWATSRALCSPPACSPFRGTISKQDVCKTLCTRQPSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVR 180
Qy 179 PFNLQVLYPLSDPALNRIHVLRDPAVRVRSRAGPILARDNGIVLGTN-GKWTEAD 237
Db 181 PFNLQSLYPLLKDPSSLNLHIVLVRDPAVRVRSRERTKGDLMIDSRIVMGQEQKJKEK 240
Qy 238 PHLRILREVCRSHVRIEAAATLKP-PFPLRGYRLVRFEDLAREPLAEIRALYFTGLTL 296
Db 241 QPYVMQVICOSQIEYK--TIQSLPKALQERYLLRYEDLARAPVATSRMYEFVGLGF 298
Qy 297 TPQLEAWIHNTHGSGIGKPIEAFHTSSRNARNYSQAWRHALPFTKILRVQEVCAQALQ 356
Db 299 LPHLQVWVHNITRCKMGD--HAFHTNARDALNYSQAWRWLSLPYKVSRLQKACCDANWL 356
Qy 357 LGYRPVTSADOQRDLTLDLVLPRGDPHFWSASPD 390
Db 357 LGYRHRVSEQQRNLLDLL-----STWTVPE 383

RESULT 8
Q81V46 PRELIMINARY; PRT; 370 AA.
ID Q81V46;
AC Q81V46;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CHST4 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22338257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC035282; AAH35282.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 48.7%; Score 999; DB 2; Length 370;
Best Local Similarity 55.9%; Pred. No. 6.4e-81;
Matches 198; Conservative 51; Mismatches 91; Indels 14; Gaps 6;

Qy 40 DRVHLVLSWRSGSSFLGQPSQHPDVFYLMPEAWHWTTLSGSAATLHMVDRDLMS 99
Db 25 ERMHVLVLSWRSGSSFVGQHPDVFYLMPEAWHWMTFKSTAWMLHMVDRDLIRA 84
Qy 100 IFLCMDVDFDAYM-PQSRNLSPFNWATSRALCSPPACSPFRGTISKQDVCKTLCTRQ 158
Db 85 VFLCDMSVDFDAYMFGPPRQSSIFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQ 144
Qy 159 FSLAREACRSYSHVVLKEVRFFNLQVLYPLSDPALNRIHVLRDPAVRVRSRAGPI 218
Db 145 FEVVEKACRSYSHVVLKEVRFFNLQSLYPLLKDPSSLNLHIVLVRDPAVRVRSRERTKGD 204
Qy 219 LARDNGIVLGTN-GKWTEADPHLRILREVCRSHVRIEAAATLKP-PFPLRGYRLVRFED 276
Db 205 LMIDSRIVMGQEQKJKEKDPQPYVMQVICOSQIEYK--TIQSLPKALQERYLLRYED 262
Qy 277 LAEPLAEIRALYFTGLTLTPQLEAWIHNTHGSGIGKPIEAFHTSSRNARNYSQAWRH 336
Db 263 LARAPVATSRMYEFVGLGFELPHLQVWVHNITRCKMGD--HAFHTNARDALNYSQAWRW 320
Qy 337 ALPFTKILRVQEVCAQALQILGYPVYSADQQRDLTLDLVLPRGDPHFWSASPD 390
Db 321 SLPYKVSRLQKACCDANWLGLYRHRVSEQQRNLLDLL-----STWTVPE 367

RESULT 9
Q9R111 PRELIMINARY; PRT; 388 AA.
ID Q9R111;
AC Q9R111;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE L-selectin ligand sulfotransferase (Chst4 protein).

```


Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 Genome Res. 10:1757-1771(2000).
 [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF131236; AAD33016.1; -
 DR EMBL; AK009113; BAB26078.1; -
 DR MGD; MGI:1349479; Chst4.
 DR GO; GO:0016021; C: integral to membrane; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 388 AA; 44694 MW; DIE9D7796DP8574D CRC64;
 Query Match 46.7%; Score 958.5; DB 2; Length 388;
 Best Local Similarity 55.6%; Pred. No. 2.9e-77;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;
 QY 42 VHVILVSSWSGSSGFLGQSFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVRLMRSIF 101
 DB 42 VHVILVSSWSGSSGFLGQSFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVRLMRSIF 101
 QY 102 LCDMDVFDAYM-POSRLNSAFFNATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFS 160
 DB 102 LCDMSVFDAYMFPKQSSIFQWQSRALCSAPVCDFFPAHEISSPKCKLGGQPF 161
 QY 161 LAREACRSYSHVILKEVRFNQLVYPLSDPALNLRIHVLVDRPRAVLSREAGPILA 220
 DB 162 MVEKACRSHGPFVILKEVRFNQLVYPLSDPALNLRIHVLVDRPRAVLSREAGPILA 221
 QY 221 RDNGVILGTNGKWEADPHLRLE-----VCRSHVRIAE-ATLKPPFPFLGRY 269
 DB 222 VDSHIVLG-----OHLTIKEEDQPYAMKICKSQVDIVKAIOTL--PEALQRY 270
 QY 270 RLVRFDLAREPLAEIRALYAFTGLTLTPQLEAWIHNIHSGIGKPKPIEAFHTSSRNARN 329
 DB 271 LFLAYEDLVRAPLAQTTLYKFVGLDFPLHLOTWVYVNRKGNQ--HAFHTNARNALN 328
 QY 330 VSOAQRHALPFTKILRVQEVYECAGALQLLGYRPPVYSAQQQDLTLDLV 376
 DB 329 VSOAQRWRLPVEKYSQLQDACCAMDLGLYQVRSQQEQGNLSLDLL 375
 RESULT 11
 Q794G9 PRELIMINARY; PRT; 483 AA.
 AC Q794G9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 25-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 26, Last annotation update)
 DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-sulfotransferase-1).
 GN Name=Gn6st-1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

TISSUE=Whole embryos;
 MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
 Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
 "Molecular cloning and characterization of an N-acetylglucosamine-6-O-sulfotransferase.";
 J. Biol. Chem. 273:22577-22583(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Uchimura K., Kadomatsu K., El-Pasakany F.M., Singer M.S., Izawa M.,
 RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.;
 "N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-Selectin Ligands and Lymphocyte Homing.";
 J. Biol. Chem. 279:35001-35008(2004).
 DR EMBL; AB011452; BAA32139.1; -
 DR EMBL; AB125058; BAD16775.1; -
 DR EMBL; AB011451; BAA32137.1; -
 DR GO; GO:0005615; C: extracellular space; TAS.
 DR GO; GO:0008146; F: sulfotransferase activity; IDA.
 DR GO; GO:0006044; P: N-acetylglucosamine metabolism; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 483 AA; 52830 MW; 831FA08P5FEFD70E CRC64;
 Query Match 32.4%; Score 665; DB 2; Length 483;
 Best Local Similarity 40.4%; Pred. No. 7.9e-51;
 Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;
 QY 37 GGDRLVHVLVSSWSGSSGFLGQSFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVRLD 96
 DB 113 GGDRLVHVLVSSWSGSSGFLGQSFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMVRLD 172
 QY 97 MRSIFLCDMDVFDAYMPO---SRNLS--AFFNATSRALCSPPACSAFPRGTISKQD--V 149
 DB 173 LSAIYRCDLSVFPOLYSPAGSGRNLTITLGFAGTNNKVCSSPLCPAYRKEVGLVDRV 232
 QY 150 CKTLCTRQPSLAREACRSYSHVILKEVRFNQLVYPLSDPALNLRIHVLVDRPRAVL 209
 DB 233 CKK-CPQRLAREECCRKYRTVVIKGVFVAVLAPLKLKDPALDLKVHLVDRPRAVA 291
 QY 210 RS-----REAGPILARD-----NGVILGT--NGKWEADPH--LRLEEV 246
 DB 292 SSRIRSHGLIRESLQVRSRDPRAHRMPPLEAGHKLAKGCGMGDPADYHALGAMVEI 351
 QY 247 CRSHVRIAEATLKPPFPFLGRYRLVRFDLAREPLAEIRALYAFTGLTLTPQLEAWIHN 306
 DB 352 CNSMAKTLQTA--LQPPDWLQGHVLYVRYEDLVGDPVTKLRVDFVGLLVSPENEQFALN 410
 QY 307 ITHGSG-IGKPIEAFHTSSRNARNVSOAQRHALPFTKILRVQEVYECAGALQLLGYRPPYSA 365
 DB 411 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFTQIKQVEEFCYQPMVILGYERNVP 467
 QY 366 DQORDTLTDLV 376
 DB 468 EEVKDLSKTL 478
 RESULT 12
 O88276 PRELIMINARY; PRT; 530 AA.
 ID O88276
 AC O88276;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE N-acetylglucosamine-6-O-sulfotransferase long form.
 GN Name=Chst2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.,
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
sulfotransferase";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; --
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase
SQ SEQUENCE 530 AA; 57814 MW; A11351B73C363EC CRC64;

Query Match 32.4%; Score 665; DB 2; Length 530;
Best Local Similarity 40.4%; Pred. No. 8.9e-51;
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGEDRVHVLVLSWRSGLGQLFSDHPDVFYLMPEAKHVVTTLSQGSAAATLHMAVRDL 96
Db 160 GGDKRLVYVFTTWRSGLSFFGELFNQNPVEFLYEPVHVWQKLYPGDVAVSLQGAARDM 219

QY 97 MRSIFLCLMDVFDAYMPQ---SRNLS--AFFNWTASRALCSPACSAFFPRGTISKOD--V 149
Db 220 LSAlyrCDLSVQLYSPAGSGRNLTLTGIFGAATNKVCCSSPLCPAYRKEVVGLVDDR 279

QY 150 CKTLCTRQPSFLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVL 209
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDAVLAFLKDPALDLKVHLVRDPAVA 338

QY 210 RS-----REAGPILARD-----NGIVLGT--NGKWVADPH-LRLIREV 246
Db 339 SSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLGAKGGMGPADYHLAGAMEVI 398

QY 247 CRSHVRIAEAAATLKPPFLRGRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIH 306
Db 399 CNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457

QY 307 ITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQEVCAQALQLLGYRPVISA 365
Db 458 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFYCPMAVLGVGVNSP 514

QY 366 DQORDLTLDLV 376
Db 515 BEVKDLSTKLL 525

RESULT 13
Q80WV3 ID Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler N.G.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL EMBL; BC051963; AAH51963.1; --
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 32.3%; Score 662; DB 2; Length 530;
Best Local Similarity 40.2%; Pred. No. 1.6e-50;
Matches 149; Conservative 65; Mismatches 121; Indels 36; Gaps 11;

QY 37 GGEDRVHVLVLSWRSGLGQLFSDHPDVFYLMPEAKHVVTTLSQGSAAATLHMAVRDL 96
Db 160 GGDKRLVYVFTTWRSGLSFFGELFNQNPVEFLYEPVHVWQKLYPGDVAVSLQGAARDM 219

QY 97 MRSIFLCLMDVFDAYMPQ---SRNLS--AFFNWTASRALCSPACSAFFPRGTISKOD--V 149
Db 220 LSAlyrCDLSVQLYSPAGSGRNLTLTGIFGAATNKVCCSSPLCPAYRKEVVGLVDDR 279

QY 150 CKTLCTRQPSFLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVL 209
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDAVLAFLKDPALDLKVHLVRDPAVA 338

QY 210 RS-----REAGPILARD-----NGIVLGT--NGKWVADPH-LRLIREV 246
Db 339 SSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLGAKGGMGPADYHLAGAMEVI 398

QY 247 CRSHVRIAEAAATLKPPFLRGRLVRFEDLAREPLAEIRALYAFTGLTLPQLEAWIH 306
Db 399 CNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457

QY 307 ITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQEVCAQALQLLGYRPVISA 365
Db 458 MTSGSGSSSKP---FVVSARNATQAANAWRTALTFOQIKQVEEFYCPMAVLGVGVNSP 514

QY 366 DQORDLTLDLV 376
Db 515 BEVKDLSTKLL 525

RESULT 14
Q9UED5 ID Q9UED5 PRELIMINARY; PRT; 483 AA.
AC Q9UED5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN Name=G6ST;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells";
RL J. Biochem. 124:670-678(1998).
RN (2)
RN SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RX Sakaguchi H., Kitagawa H., Sugahara K.,
RA "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence";
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -
DR EMBL; AB021124; BAB16886.1; -
DR EMBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
RW Transferase.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 32.2%; Score 659.5; DB 2; Length 483;
Best Local Similarity 39.1%; Pred. No. 2.4e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

Qy 31 GPSSPAG-----CEDRVHVLVLSWSRSGSFLGOLFQHPDVFYLMPEPAHVVMTTL 81
Db 98 GVAAPPNGTRGTGGVGDKQQLVYFTTWRSGSFFGELFNQNPVEVFLYEPVHVWQKL 157

Qy 82 SQGSAATHMAVRDLMSRIFLCMDVFDAYMPQ---SRNLS--AFFNWSRALCSPAC 136
Db 158 YPGDAVSLQGAARDMSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 217

Qy 137 SAFPRGTISKOD--VCKTLCTROPFLAREACRSYSHVVLKEVFFNLQVLYPLSDPAL 194
Db 218 PAYRKEVVLVDVDRCKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLRDPAL 276

Qy 195 NLRVHLVRDPAVLRS-----REAAGPILARD-----NGVLGTNGKWV 234
Db 277 DLKVIHLVRDPAVASSRIRSRHGLIRSLQVRSRDPRAHRMPPLEAAGHKLGAKKEGV 336

Qy 235 --EADPH-LRLIREVCRSHVRIAEATLKPPFFLRGRYLRVRFEDLAREPLAEIRALYAF 291
Db 337 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYRVYEDLVGDPVKTLRRVYDF 395

Qy 292 TGLTLTPOLEAWIHNITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQVQC 350
Db 396 VGLLVSPMEQFALNMTSGSSSKP-----FVVSARNATQAAANWRTALTFTQOIQVBEFC 452

Qy 351 AGALQLLGYRPVYSADOQDRLTLDIV 376
Db 453 YQPMNAVGLYERVNSPEEVKDLSTLL 478

RESULT 15

Q9Y4C5

ID Q9Y4C5 PRELIMINARY; PRT; 530 AA.

AC Q9Y4C5; Q9GZNS; Q9Y6F2;

DT 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)

DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)

DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)

DE (Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells";
RL J. Biochem. 124:670-678(1998).
RN (2)
RN SEQUENCE FROM N.A.
RP TISSUE=Umbilical vein endothelium;
RC MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RX Li X., Tedder T.F.;
RA "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization";
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -
DR EMBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
RW Transferase.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match 32.2%; Score 659.5; DB 2; Length 530;
Best Local Similarity 39.1%; Pred. No. 2.8e-50;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

Qy 31 GPSSPAG-----CEDRVHVLVLSWSRSGSFLGOLFQHPDVFYLMPEPAHVVMTTL 81
Db 145 GVAAPPNGTRGTGGVGDKQQLVYFTTWRSGSFFGELFNQNPVEVFLYEPVHVWQKL 204

Qy 82 SQGSAATHMAVRDLMSRIFLCMDVFDAYMPQ---SRNLS--AFFNWSRALCSPAC 136
Db 205 YPGDAVSLQGAARDMSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 264

Qy 137 SAFPRGTISKOD--VCKTLCTROPFLAREACRSYSHVVLKEVFFNLQVLYPLSDPAL 194
Db 265 PAYRKEVVLVDVDRCKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLRDPAL 323

Qy 195 NLRVHLVRDPAVLRS-----REAAGPILARD-----NGVLGTNGKWV 234
Db 324 DLKVIHLVRDPAVASSRIRSRHGLIRSLQVRSRDPRAHRMPPLEAAGHKLGAKKEGV 383

Qy 235 --EADPH-LRLIREVCRSHVRIAEATLKPPFFLRGRYLRVRFEDLAREPLAEIRALYAF 291
Db 384 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYRVYEDLVGDPVKTLRRVYDF 442

Qy 292 TGLTLTPOLEAWIHNITHGSG-IGKPIEAFTSSRNARNVSOAQRHALPFTKILRVQVQC 350
Db 443 VGLLVSPMEQFALNMTSGSSSKP-----FVVSARNATQAAANWRTALTFTQOIQVBEFC 499

Qy 351 AGALQLLGYRPVYSADOQDRLTLDIV 376
Db 500 YQPMNAVGLYERVNSPEEVKDLSTLL 525

Search completed: June 23, 2005, 08:48:58

Job time : 43.7985 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.8459 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLFSSKTVTVLLLAQTT.....LTLDLVLPRGDFHSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_79: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	659.5	32.2	484	2 JE0261	N-acetylglucosamin
2	613.5	29.9	484	2 JC7350	N-acetylglucosamin
3	610	29.7	486	2 JC7351	N-acetylglucosamin
4	540.5	26.4	458	2 A57397	chondroitin 6-sulf
5	107.5	5.2	307	2 E95934	probable enzyme, C
6	105	5.1	359	2 T16350	hypothetical prote
7	102.5	5.0	388	2 G70729	hypothetical prote
8	96	4.7	388	1 DEHUP1	pyruvate dehydroge
9	96	4.7	707	2 S29029	phenylalanine ammo
10	95	4.6	474	2 T38905	probable valine-py
11	95	4.6	885	1 S78246	endopeptidase Clp
12	94.5	4.6	1581	1 VGMJBV	peptomer glycoprot
13	94	4.6	582	2 A70841	probable oxalyl-Co
14	93.5	4.6	615	2 C75278	DNA polymerase III
15	93	4.5	523	2 B83629	probable ATP-bindi
16	90.5	4.4	380	2 T42755	tyrosylprotein sul
17	88	4.3	737	2 AD2037	hypothetical prote
18	86.5	4.2	445	2 T05639	hypothetical prote
19	86.5	4.2	577	2 B75595	probable long-chai
20	86.5	4.2	712	2 T32331	hypothetical prote
21	86	4.2	417	2 F71238	hypothetical prote
22	85.5	4.2	221	2 T50665	hypothetical prote
23	85.5	4.2	389	1 DSPGPA	pyruvate dehydroge
24	85.5	4.2	425	2 E83023	3-deoxy-D-manno-Oc
25	85.5	4.2	571	2 S69210	protein kinase cak
26	85.5	4.2	1254	2 T41262	mutS family DNA mi
27	85	4.1	364	2 A49781	cholesterol dehydr
28	85	4.1	696	2 A11029	hypothetical prote
29	85	4.1	776	2 T31210	membrane protein t

30	85	4.1	1182	2 S14916	hypothetical prote
31	84.5	4.1	334	2 T44163	hypothetical prote
32	84.5	4.1	390	1 DERTP1	pyruvate dehydroge
33	84.5	4.1	390	1 DERTPA	pyruvate dehydroge
34	84.5	4.1	390	2 S23506	pyruvate dehydroge
35	84.5	4.1	431	2 A36036	cytochrome P450 2F
36	84.5	4.1	1844	1 RRWFTM	genome polyprotein
37	84	4.1	552	2 T49254	flagella-related p
38	84	4.1	590	2 T39246	protein phosphotas
39	83.5	4.1	260	2 B95965	probable two-compo
40	83.5	4.1	390	1 DERUPA	pyruvate dehydroge
41	83.5	4.1	420	2 A53531	oncofetal trophobl
42	83.5	4.1	856	2 F87316	conserved hypothet
43	83.5	4.1	910	2 E89918	2-oxoglutarate deh
44	83.5	4.1	1844	2 S01956	hypothetical prote
45	83.5	4.1	1985	2 S19151	hypothetical prote

ALIGNMENTS

RESULT 1

JE0261

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human

C:Species: Homo sapiens (man)

C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002

C:Accession: JE0261

R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998

A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of (

A:Reference number: JE0261; MUID:98391845; PMID:9722682

A:Accession: JE0261

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679

C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine to 6-sulfotransferase

C:Superfamily: chondroitin 6-sulfotransferase

C:Keywords: sulfotransferase

Query Match 32.2%; Score 659.5; DB 2; Length 484;
Best Local Similarity 40.0%; Pred. No. 1.5e-51;
Matches 150; Conservative 67; Mismatches 121; Indels 37; Gaps 12;

Qy	34	SPAGGEDRVH-VLVLSSWRSSPLGOLFQHPDVFVLMPEAHVWTTLSGSAATHMA	92
Db	110	APEGVGDKRHMMYVFTTWRSQSGFFGELFNQNPVEVFLYEPVHWQKLYPGDAVSLQA	169
Qy	93	VRDLMRSFILCDMDVFDAYMPQ---SRNLS--AFFNWTSRALCSPACSAFPRGTISKQ	147
Db	170	ARDMLSALYKCDLSVFQLYSPAGSGGNLTTLTGIFGATNKVVCSSPLCPAYRKEVWGLV	229
Qy	148	D--VKTKLCTQRPFSLAREACRSYSHVVLKEVRFNFNLOVLYPLSLDSPALNURIHVLRDP	205
Db	230	DDRVCCK-CPQRLARPEECRKYRTLVIKGVRFVDAVLAPLRDPAALDKVHLVRDP	288
Qy	206	RAVLRS-----REAAGPILARD-----NGVLGTNGKWV--EADPH-LRL	242
Db	289	RAVASSIRSRHGLIRESLQVRSRDPRAHMPFLEAGHKLGAKGGVGGPADTHAGA	348
Qy	243	IREVCRSHVRIAEATLKPPFLRGVRLRVFEDLAREPLAEIRALYAFTLTLPQLEA	302
Db	349	MEVICNSMAKTLQTA-LQPPDWLGCHLVVRYEDLVGEVPTKLRVDFVGLLSPENEQ	407
Qy	303	WIHNITHGSG-IGKPIBAFTSSRNARNVSOAMRHAPFTTKILRVQEVACAGALQLLVPR	361
Db	408	FALNMTSGSGSSSKP---FVVSARNATQAAANWRTALTFFQIKQVEEFQYQPMVAVLYER	464
Qy	362	VYSADQQRDITLDLV	376
Db	465	VNSPEEVKDLKSTLL	479

RESULT 2
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C;Accession: JC7350
R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7350
A;Molecule type: mRNA
A;Residues: 1-484 <UCH>
A;Cross-references: UNIPROT:Q99NB0; UNIPROT:Q9EP78; DDBJ:AB040710
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.9%; Score 613.5; DB 2; Length 484;
Best Local Similarity 38.8%; Pred. No. 2.1e-47;
Matches 146; Conservative 54; Mismatches 127; Indels 49; Gaps 11;

QY 30 PGSPSPAG---GSDRVHVLSSWRSGSFLGOLFQSDHPDVFYLMPEPAWHVMTTSLSQGS 86
DB 85 PGNLSAVGEAVTQEKQHIYVHATWRTGSSFLGELFNQHPDVFYLYSPMWHLWQALYPGDA 144

QY 87 ATLHMVDRMLRSIFICDMDFDAY-----MQSRNL--FPFNWATSRALCSPPA 135
DB 145 ESQGLALDMLRSLFRCDVSVLRVYAPGDPGGERAPDSANLTATMLFRWRTNKVICSPPL 204

QY 136 CSAFPRGT---TSKQDVCKTCTROPFSL-AREA-CRSYSHVVLKEVRFNLOVLYPLL 189
DB 205 CPAAPRARADVGLVEDKACSTC--PPVSLRALEAECKYPVVVIKDVRLDGLVLPVLL 262

QY 190 SDPALNLRIVHLVRDPAVRLSREAAAGPILARDNGVLGTNGK-----WVEADP 238
DB 263 RDPGLNKKVQLFRDPRAVNSRLKSRQGLRESIQVLRTRQGRDHFHVRVLLAHGVDPAR 322

QY 239 --HLRLEVCRSHVRIAEAA-----TLKPPPLFGRVLRVRFEDLAREPLA 293
DB 323 GGOARALPSAPRADFFLTSALEICEAWLRDLLFTRGAPAWLRRLRYEDLVVQPOA 382

QY 284 EIPALYFTGLTLPOLAEIHNITGSGIGKPIEAFTSSRNARNVQAWRHALPFTKI 343
DB 383 QLRRLRFRSGRLTALDAFAFNTRGSAYGAD-RPFHLSARDAREAVHWRLRSLQEQV 441

QY 344 LRVQEVCAQALQLLGY 359
DB 442 RQVETACAPAMRLAY 457

RESULT 3
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C;Species: Homo sapiens (man)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C;Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7351
A;Molecule type: mRNA
A;Residues: 1-486 <UCH>
A;Cross-references: UNIPROT:O75667; DDBJ:AB040711
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.7%; Score 610; DB 2; Length 486;
Best Local Similarity 38.8%; Pred. No. 4.4e-47;
Matches 147; Conservative 52; Mismatches 126; Indels 54; Gaps 11;

QY 30 PGSPSPAG---DRVHVLSSWRSGSFLGOLFQSDHPDVFYLMPEPAWHVMTTSLSQGS 85
DB 86 PNLNSAVGEAVSRKQHIYVHATWRTGSSFLGELFNQHPDVFYLYSPMWHLWQALYPGD 145

QY 86 AATLHMVDRMLRSIFICDMDFDAY-----MQSRNL--SAFFNWATSRALCSPSP 134
DB 146 AESLQGLALDMLRSLFRCDVSVLRVYAPGDPAPRADPTANLTATLFRWRTNKVICSP 205

QY 135 ACSAPRGITISQDVCKTCTCTQ--PFSL-AREA-CRSYSHVVLKEVRFNLOVLYPLL 190
DB 206 LCPGAPRARAEVGLVEDTACRSCTPVAIRALEAECKYPVVVIKDVRLDGLVLPVLLR 265

QY 191 DPALNLRIVHLVRDPA-----VLSREAAAG-----PILARDNGVLVG 228
DB 266 DPGNLNKKVQLFRDPRAVNSRLKSRQGLRESIQVLRTRQGRDHFHVRVLLAHGVGARP 325

QY 229 TNGKWEADPH-----LRLIREVCRSHVRIAEAAATLKPPPLFGRVLRVRFEDLARE 280
DB 326 GQSRALPAAPRADFFLTGALEICEAWLRDLLFARGA----PAWLRRRYLRVLRVEDLV 381

QY 281 PLAETRALYFTGLTLPOLAEIHNITGSGIGKPIEAFTSSRNARNVQAWRHALPFP 340
DB 382 PRAQLRRLRFRSGRLAALDAFALNMTGRGAAYGAD-RPFHLSARDAREAVHAWRRLSR 440

QY 341 TKILRVOEVCAGALQLLGY 359
DB 441 EQVRQVEAACAPAMRLAY 459

RESULT 4
AS7397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C;Accession: AS7397
R;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kinata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A;Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransfe
A;Reference number: AS7397; MUID:95355490; PMID:7629189
A;Accession: AS7397
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-458 <FUK>
A;Cross-references: UNIPROT:Q92179; GB:D49915; NID:9971262; PIDN:BAA08655.1; PID:9971263
C;Superfamily: chondroitin 6-sulfotransferase
C;Keywords: sulfotransferase

Query Match 26.4%; Score 540.5; DB 2; Length 458;
Best Local Similarity 34.3%; Pred. No. 7.9e-41;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

QY 36 AGGEDRVHVLSSWRSGSFLGOLFQSDHPDVFYLMPEPAWHV--WTTLSQGSAAHLMAV 93
DB 107 AAPEPRHRVLLMATRTGSSFVGEFFNQGNIFYLFEPLWHIERTVTFFPGGANAVGSAL 166

QY 94 --RDLNRSIFICDMDFDAYM---PQSRNLSAFFNWATSRALCSPSPACSAFPRGTISKQD 148
DB 167 VYRDVLQQLLDCDLYLSEFISPAPEEHTALFRGSGSHSLCEEPVCTPSLKKVFEKYH 226

QY 149 VCKTLCTROPFSLAREACRSYSHVVLKEVRFNLOVLYPLLSDPALNLRIVHLVRDPAV 208
DB 227 CKNRRCGPLNITLAAEACRRKQHMALKTVRIQLFQLPLAEDPRDLAIQLVRDPAV 286

QY 209 LRSREAAAGPILARDNGVLGTNGKWE-----ADPHLRIRVCRSHVRIAEAAATL 259
DB 287 LVSRMVA-----PSGKYESWKWAAEAPLQEDDEVQRLRGNCES-IRLSAELGL 335

QY 260 KPFPFLGRVLRVRFEDLAREPLAEIRALYATGLTLPOLAEIHNITGSGIGKPIEA 319
DB 336 RQPRWLGRYMLRVYEDVARPLRKALEYRFAGHFTPOVEEWIRANTQAP---QDSNG 392

QY 320 FHTSSRNARNVQAWRHALPFTKILRVQEVCAQALQLLGYRVPVYSADQQORDLTLDLVLPR 379

Db 393 IYSTQKNSQSEKWRFSIPFKLQAVQVQDACEPAMRLFGYKLASSAQLNTRSL-LEE 451
 Qy 360 GP 381
 Db 452 GP 453

RESULT 5
 E95934
 Probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: T16350
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95934
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>
 A:Cross-references: UNIPROT:Q92V64; GB:AL591985; PIDN:CAC49141.1; PID:gl5140626; GSPDB:C
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SM521237
 A:Genome: plasmid

Query Match 5.2%; Score 107.5; DB 2; Length 307;
 Best Local Similarity 20.6%; Pred. No. 0.064;
 Matches 78; Conservative 46; Mismatches 131; Indels 123; Gaps 17;

Qy 32 PSSPAGGEDRVHLVLSW-RSGSFLGQLFSQHPDVF---YLMPEPAWHVMTTILSQSAA 87
 Db 2 PSQP-----VRIAYIAGVGRSGTILDIALQHAHVAGAEITSLTRHV---RHNEVC 52
 Qy 88 TLHMAVRLMRSIFLCD-----MDVFDAYMPQSRNLSAFENWATSRAL 130
 Db 53 ACQNAIRD-----CSFWSSVVRREWSGDQPLMEYCALQOKFEGLSMMTRLSSGMGL 105
 Qy 131 CSPPACSAFPGRTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVFFNQLVLYPLLS 190
 Db 106 -----GKQSLYLHTKRLFS-AMQSC-SGRQVIVDSKKLPGRAMAVAQI- 148
 Qy 191 DPALNLRVHLVRDPAVRLSREAGPILARDNGIVLGTNGKWKVADPHLRILREVCRSH 250
 Db 149 -PGIDMRVHLVRDGRGV-----AWSLL-----KGYERDAKSLQKEI---- 185
 Qy 251 VRIABAAATLKPPPLRGYR-----LVRFEDLAREPLAIRALY 289
 Db 186 -----KPKSVFRTALRSMWNLAVEYLSRKLGSKWKVRVYEDFASDPVAVMOQIG 236
 Qy 290 AFTGLTLTPOLEAWTHNTHGSGIGKPTAEFTSSRN-----ARNVSOAWRHLPPT 341
 Db 237 TFLDLDS-QVGTSEN-----GEAMGPGHQVAGNRLRMWASIALNKDETWRTMRPAR 288

Qy 342 KILRVQEVACAGALQLGY 359
 Db 289 QQVVSFQRLGGWMLRYGY 306

RESULT 6
 T16350
 hypothetical protein F42G9.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T16350
 R:Taich, A.
 submitted to the EMBL Data Library, March 1996
 A:Description: The sequence of C. elegans cosmid F42G9.
 A:Reference number: 218498
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <TAI>
 A:Cross-references: UNIPROT:Q20351; EMBL:U00051; NID:gl21216308; PID:AAA913;
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:F42G9.8
 A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
 C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;
 Best Local Similarity 19.8%; Pred. No. 0.13;
 Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

Qy 11 VTVLLLAQTTCLLLF1-----ISRPGSPSPAGGEDRVH-----VLV 46
 Db 31 IYIFIFCTTCLLIFSSIKCKKLEQLESLKESLIFNEQDARHSRRLLSNLEQLIFV 90
 Qy 47 LSSWRSGSFLGQLFSQHPDV-----FYLMPEPAW-HVMTTILSQSAAATLHMAVR 94
 Db 91 GGVPRSGTTLRAILDADHPDVRCCGCTMTLPSPLTWAGWENDV---NNSGIT----- 141
 Qy 95 DLMSRIFLCMDVDYAYMPQSRNLSAFENWATSRALCSPPACSAFPGRTISK-QDVCKTL 153
 Db 142 -----QEVFD-----DAVSAPITEIVAKHSELAPRL 167
 Qy 154 CTROPFSLAREACRSYSHVVLKEVFFNQLVLYPLLSDPALNLRVHLVRDPAVRLSR- 212
 Db 168 CNKDP-----YTALMLPTIR-----RLYP-----NAKFLMIRARAVVHSMI 205
 Qy 213 EAAGPILARDNGIVLGTNGKWKVADPHLRILREVCRSHVRIAEATLKPPPLRGYRLV 272
 Db 206 ERKVPVAGYNTSDISMPQW---NQELRKMTFCNN---APQCIC-----V 247
 Qy 273 RFEDLAREPLAIRALYAFGLTLTPOLEAWHNTGSGIGKPI-----EAPHTSS-RNA 327
 Db 248 YVERLIQKPAEILRITNFDLPFSQW-----LRHQDLICDEVLDNDQSFASQVKNS 301
 Qy 328 RNVS--QAWRHLPPTKILRVQEVACAGALQLGY-----RPVYS 364
 Db 302 INKALTSWDFCFSEETLRKLDV-APPLGLTGYDTSISKPDYS 344

RESULT 7
 G70729
 Hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C:Accession: G70729
 R:Coie, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: G70729
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-388 <COL>
 A:Cross-references: UNIPROT:Q50695; GB:Z77163; NID:gl2123456; PID:G3261610; PIDN:CAB00968.1
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv2267c

Db 128 NIIMTGNSTCLDIALRMLNTRGDSILVEKVSFSPALQSMRPLGLSCIPIDMDQF-GFLP 186
 Qy 114 QSRNLSAFFNW-ATSRALCSPACSAFPRGTISKQDVCKTCTRPQPSLAREACRSYSHV 172
 Db 187 ESMD-DILTNDATSYGSPKPHVLYTPTGQNPSTLSVERRKQIYTLAQK-----HDII 241
 Qy 173 VLKVRFNQV-LY-----PLSDPALNLRIVHLVRDPAVLRSEAGPILARN-----GI 225
 Db 242 ILEDEPYVYLQMDAYEGKPEAAKRAFTNE--QFKELIPSPFLSDVGRVIRMDSLSKV 299
 Qy 226 VLGTNGKWEADPHLRILIREVCRSHVIRAEATLKPPFLRG-RYLRVP--EDLAREPL 282
 Db 300 AGRSRVGNFTAQPLF-----IERGLAAETATQASGISQGIYAMPKHWGQGYLEWL 353
 Qy 283 AEIRALYAF 292
 Db 354 KHIR--YSYT 361
 RESULT 11
 S78246
 N:Altepidase Clp (EC 3.4.21.-) ATP-binding chain clpC [similarity] - Odontella sinensis
 N:Altepidase names: ATP-dependent Clp proteinase regulatory chain; caseinolytic Clp prote
 C:Species: adenosinetriphosphatase (EC 3.6.1.3)
 C:Species: chloroplast Odontella sinensis
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
 C:Accession: S78246
 P:Kowalik, K.V.; Stoebe, B.; Schaffran, I.; Kroth-Pancic, P.; Freier, U.
 Plant Mol. Biol. Rep. 13, 336-342, 1995
 A:Title: The Chloroplast Genome of a chlorophyll a+c containing Alga, Odontella sinensis
 A:Reference number: S78238
 A:Accession: S78246
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-885 <KOW>
 A:Cross-references: UNIPROT:P49574; EMBL:Z67753; NID:g1185127; PIDN:CAA91619.1; PID:g118
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
 C:Genetics:
 A:Gene: clpC
 A:Genome: chloroplast
 C:Function:
 A:Description: allows clpP to hydrolyze polypeptides and proteins, probably by a chaperon
 e activity; ATP hydrolysis is required for Clp binding of proteins but not of smaller
 C:Superfamily: endopeptidase Clp ATP-binding chain
 C:Keywords: ATP; chloroplast; hydrolase; molecular chaperone; nucleotide binding; P-loop
 F:218-225/Region: nucleotide-binding motif A (P-loop)
 F:285-290/Region: nucleotide-binding motif B
 F:560-567/Region: nucleotide-binding motif A (P-loop)
 F:628-633/Region: nucleotide-binding motif B
 F:224/Binding site: ATP (Lys) #status predicted
 F:566/Binding site: ATP (Lys) #status predicted
 Query Match 4.6%; Score 95; DB 1; Length 885;
 Best Local Similarity 21.0%; Pred. NO. 3.5;
 Matches 89; Conservative 59; Mismatches 161; Indels 114; Gaps 18;
 Qy 5 RPSKTVTVLLAQ-----TTCLALFII-SRPGSSPAGGEDRVHVLVLSWR 51
 Db 4 KFTGAIVMLSQBEARMGNHFNVTGTEQLLGIIGQRHGIGARALKKQV---TLKKAR 60
 Qy 52 -----SGSFLGQLFSQHPDVFYLMPEAWHVTTLTSSQAATLHMAVRDLMS--IF 101
 Db 61 REIELYIGRTGTFVASEIPPTPRAKRVLEMAVHEGKDLQGNFVTEHILLALISESDGA 120
 Qy 102 LCDMDVFDAYMQPSNLSAFTNATSRALCSPACSAFPRGTISKQDVCKTCTRPQPSL 161
 Db 121 MRTLKLGWNPVKLRNLILMWI-----EENQEEILRPLTQAEKPL 161
 Qy 162 AREACRS-----YSHVVLKVRFNQVLYPLSDPALNL-RIVHLVRDPAVLRSE 213
 Db 162 EREKKSSTPTLDEYSENISEAVDGKL-----DPVIGRDKIEHEV-----IKVLARR 210
 Qy 214 AAGPILARDNGIVLGTNGKWEADPHLRILIREVCRSHVIRAEATLKPPFLRGYRLVR 273

Db 211 KNPVLIGEGV-----GKTAVAEGLAQLI-----IAE-----KAPFLDGNL----- 248
 Qy 274 FEDLAREPLAEIRALYAFGLTLPQLEAWIHNITHSGIGKPIEAFT-----SSRNAR 328
 Db 249 ---LMALDLSILAGTKYRG-EFEERIKRIVEEVQNDASAILVIDEITLTVGAGAEGAV 304
 Qy 329 NVSQARHALPFTKILRVQVVCAGALQLLYR-----PYVSADQQRDLTLD 374
 Db 305 DAANILKPALARGKFR-----CIGATTIDYKRYIERDPALERFPQVHVKEPTVGVTIE 359
 Qy 375 LVL 377
 Db 360 ILL 362
 RESULT 12
 VGMJBV
 peplomer glycoprotein precursor - Berne virus (strain P138/72)
 C:Species: Berne virus
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
 C:Accession: A36759
 P:Snijder, E.J.; Den Boon, J.A.; Spaan, W.J.M.; Weiss, M.; Horzinek, M.C.
 Virology 178, 355-363, 1990
 A:Title: Primary structure and post-translational processing of the Berne virus peplomer
 A:Reference number: A36759; MUID:91020973; PMID:2219698
 A:Accession: A36759
 A:Molecule type: genomic RNA
 A:Residues: 1-1581 <SNI>
 A:Cross-references: UNIPROT:P23052; GB:X52506; NID:g62059; PIDN:CAA36748.1; PID:g62060
 C:Genetics:
 A:Gene: P
 C:Superfamily: Berne virus peplomer glycoprotein
 C:Keywords: glycoprotein; transmembrane protein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1581/Product: peplomer glycoprotein #status predicted <PGP>
 F:1547-1572/Domain: transmembrane #status predicted <TMN>
 F:25,310,384,494,574,935,969,1267,1327,1385,1389,1428,1431,1438,1483,1487,1495,1515
 Query Match 4.6%; Score 94.5; DB 1; Length 1581;
 Best Local Similarity 17.6%; Pred. No. 8.3; Mismatches 65; Indels 179; Gaps 19;
 Matches 81; Conservative 81; Mismatches 65; Indels 179; Gaps 19;
 Qy 2 WLPFRFSKTVTVLLAQTTCLLPIISRPSPAGGEDRVHVLVLSWRSSGFLGQLF 61
 Db 259 WYVAFQNKATAVILPSELIVPAQKVR-----RLG 289
 Qy 62 SQHPDVFYLMPEAWHVTTLTSSQAATLHMAVRDLMSIFLCDMDVFDAYMQPSNLSAF 121
 Db 290 VNTPDYFVLVKQAVH-----YLSQA-NLSP- 313
 Qy 122 FNWATSRALCSPACSAFPRGTISKQDVCKTCTRPQPSLAREACRSYSHV-LKEVRF 180
 Db 314 -NYALFALCN-----SLYQSSATLSTLCGSPFFVAQEC---YNNALYLPDAVFT 360
 Qy 181 NL-----QVLYPL-----LSDPALNL-----RIVHLVRDPAVL----- 209
 Db 361 TLESTLSWDYQINFLNQVLTQNETFLQLPATNYQGTLSQGRMLNLFKDAIVLDFD 420
 Qy 210 -----RSREAGP-----ILARDNGIVLGTNGKWEADPHLRILIREVCRSHVIRAEATLKP 261
 Db 421 TKFVRTNDAPSSDIFVYVVARQAQLIRYGNFRIEQINGYFQV---KCSSNI-----ISTLEP 473
 Qy 262 PFLRGYRLVRFPEDLAREPLAEIRALYAF----- 292
 Db 474 HP---AGVIMIRARHSHMWSVAARNSTSFYCVTHSTLTTFGKLDISTSWFFHTLALSPGPVSQ 531
 Qy 293 -----GLTLPQLEAWIHNIT-----HGSGIGKPIEAFTSSRNARVS 331
 Db 532 VSMPLLSAAVGVYMHPIEHWIPLLLLAQSQYQPSFFNIGINKTI-TLTTQLQAYAVY 590
 Qy 332 QAWRHALPFTKILRVQVVCAGALQLLYRVPYVSADQQRDL 371

Db 591 TAMFLSVYVRLPEARLTLG-VOLVFFIQALLSIKQADL 629

RESULT 13

A70841

probable oxalyl-CoA decarboxylase - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 16-Aug-2004

C:Accession: A70841

R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70841

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-582 <COL>

A:Cross-references: UNIPROT:O53639; GB:AL021926; GB:AL123456; NID:G3261520; PIDN:CAA1731

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: oxaA

C:Superfamily: Acetolactate synthase, large subunit/pyruvate oxidase; thiamin pyrophosph

Query Match 4.6%; Score 94; DB 2; Length 582;

Best Local Similarity 21.4%; Pred.No.2.5; Mismatches 44; Indels 60; Gaps 11;

Matches 55; Conservative 44; Mismatches 98; Indels 60; Gaps 11;

QY 152 TCTROPFSLAREACRSYSHVVLKEVFFNQLVLYPLSDP-----ALNLRIVHLV 202

Db 2 TTRSASCTVLTDC-----HLVDALKANDVITYGVGIPITDLARAAQASGIRYIGF- 56

QY 203 RDPFAVLRSRAGPILARONGIVLTNGK-WFADPHLRLIREVCRSHVRIABAATLKP 261

Db 57 RHEASAGNAAAAGFLTARP-GVCLTTSGPGLNGLPALANATTNCFPMIQTISGSSS-RP 114

QY 262 -----PPFLGRYLVRFEDLAREPLAEIRALYAFVLTGLTLPQLEAW 303

Db 115 MVLQRGYDQLQNLNARFPVKAARYIGQVDIGRGVARIR-----TATSRPGG 166

QY 304 IHNITHSGIGKPIEAPHTSSRNARNVSAQRHALPFTKILRVQEVGAGALQLLGYRPVY 363

Db 167 VYLDIPGVLGQAVEASAAS-----GAIWRPVDPAPELLPAPEIDRALDVL----- 213

QY 364 SAQQORDLTLDLVLPRG 380

Db 214 -AQQRPL---LVLSKG 226

RESULT 14

D75278

DNA polymerase III, tau/gamma subunit - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: C75278

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: C75278

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-615 <WHI>

A:Cross-references: UNIPROT:Q9RRS5; GB:AE002071; GB:AE000513; NID:G6460218; PIDN:AAF1195

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2410

A:Map position: 1

C:Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 4.6%; Score 93.5; DB 2; Length 615;

Best Local Similarity 23.4%; Pred.No.2.9;

Matches 92; Conservative 41; Mismatches 141; Indels 119; Gaps 21;

QY 4 PRPSSKTVTVLLAQTCTLLLFIIISRPSPAGGEDRVHVLVLSWSRSGSSFLGQLFSQ 63

Db 44 PRGVGKTTTARLIAMTA-----NCTGPAPKPGCECECLAV-----RAGS----- 83

QY 64 HDVDFVLMPEAMHVWTTLSQGSAAHLMA--VRDLMSIFLDCMD-----VFDAYMPQS 115

Db 84 HPDVMIEI-----DAASNNVDVDRLEKVGKLAAMRGKKIYIIDEAHMMS 129

QY 116 RNLSAFENWATSRALCSPSPACAPPRGTISKQDVCKTCTROPFSLAREACRSY--SHV 173

Db 130 R---AAFN-ALLKTLSEPEHVFILATTEPEKIIITLSR-----COHYRFRRLT 176

QY 174 LKEV--RFFNLQVLYPLSDP-ALNL--RIVH-LVRDPRAVLRSRAAGPILARDNGIVL 227

Db 177 SEEIACKLAGLVLTLEGASADPDALNLIGRLADGAMRDGESSLLERMLAAGTAVTRP----- 231

QY 228 GTNGKWEADPHLRLIREVCRSHVRIAEATLKPPFPLRGYLVRFEDLAREPLAEIRA 287

Db 232 -----AVEEALGUPPGERVGVASALLVGD--AGEAISGAQA 266

QY 288 LY--AFTGLTLTPQLEAWIHNITHGS-GIGKPIEAPHTSSRNARNVSAQRHALPFTKIL 344

Db 267 LYRDGFAARTVVEGLVAFAALHABLGGE-----EGRLEGAEVPRLL 310

QY 345 RVQEVGAGALQLLGYRPVYSADQORDLTLDLVL 377

Db 311 KLQ---AALDEQEARFARSADQO---SLELAL 336

RESULT 15

B83629

probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aerugin

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C:Accession: B83629

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brj adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, ; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathox

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: B83629

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-523 <STO>

A:Cross-references: UNIPROT:Q91626; GB:AE004451; GB:AE004598; NID:G9945958; PIDN:AAG0352

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA0136

Query Match 4.5%; Score 93; DB 2; Length 523;

Best Local Similarity 26.4%; Pred.No.2.6;

Matches 37; Conservative 17; Mismatches 60; Indels 26; Gaps 5;

QY 13 VLLLAQTCTLLLFIIISRPSPAGGEDRVHVLVLSWSRSGSSFLGQLFSQHPDVFYLM 72

Db 384 VLALAE-EIIRRFAPKAPGAGAPARS-----LSGGNLQKFILGREILQAPRLVAAH 434

QY 73 PAHWVWTTLSQGSAAHL--NAVRDLMSIFLDCMDVFDAYMPQSRNLSAPFNWATSR 129

Db 435 PTWGV----DVGAALIHRLIALRDAGTAVLVVSEDLDELFLLSDR-----IAA 480

QY 130 LCSPPACAPPRGTISKQDV 149

Db 481 LCSGRCLCPAVATASAPQKV 500

Search completed: June 23, 2005, 08:50:12

us-10-697-828-8.rpr

✓ Thu Jun 23 12:39:05 2005

Job time : 9.8459 secs

THIS PAGE BLANK (USPTO)

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2057	100.0	395	2	Q90UP4	Q9gup4	mus musculus
2	1708	83.0	395	2	Q9GZX3	Q9gzx3	homo sapien
3	1536.5	74.7	390	2	Q71CN3	Q71cn3	homo sapien
4	1536.5	74.7	411	2	Q9GZS9	Q9gzs9	homo sapien
5	1187	57.7	392	2	Q6GNC39	Q6gnc39	xenopus lae
6	1022.5	49.7	386	2	Q8NCG5	Q8ncg5	homo sapien
7	1017.5	49.5	386	2	Q9Y5R3	Q9y5r3	homo sapien
8	1010.5	49.1	370	2	Q81V46	Q81v46	homo sapien
9	991	48.2	388	2	Q9R111	Q9r111	mus musculus
10	984	47.8	388	2	Q9WUE5	Q9wue5	mus musculus
11	640	31.1	483	2	Q794G9	Q794g9	mus musculus
12	640	31.1	530	2	Q88276	Q88276	mus musculus
13	637	31.0	530	2	Q80WV3	Q80wv3	mus musculus
14	636.5	30.9	483	2	Q9UED5	Q9ued5	homo sapien
15	636.5	30.9	530	2	Q9Y4C5	Q9y4c5	homo sapien
16	585	28.4	420	2	Q6DBY9	Q6dby9	brachydanio
17	581	28.2	484	2	Q9BP78	Q9bp78	m n-acetyl
18	580	28.2	486	2	Q75667	Q75667	homo sapien
19	579	28.1	486	2	Q9NS84	Q9ns84	homo sapien
20	578	28.1	484	2	Q9N8B0	Q99nb0	mus musculus
21	576	28.0	485	2	Q6XQG8	Q6xqg8	rattus norv
22	563.5	27.4	479	2	Q71GC8	Q71gc8	homo sapien
23	561.5	27.3	479	2	Q75099	Q75099	homo sapien
24	551	26.8	411	2	Q9EQC0	Q9eqc0	mus musculus
25	550	26.7	472	2	Q88199	Q88199	mus musculus
26	549	26.7	411	2	Q43916	Q43916	homo sapien
27	542	26.3	474	2	Q9QZL2	Q9qz12	rattus norv
28	495	24.1	458	1	C6S7_CHICK	Q92179	gallus gall
29	457.5	22.2	441	2	Q93403	Q93403	torpedo cal
30	373.5	18.2	257	2	Q79415	Q79415	mus musculus
31	320.5	15.6	225	2	Q6RY62	Q6ry62	cavia porce

```

QY 241 VVNEVCSSHVRVIAEAAHLKPPFLQDRYRLRVYVEDLARDPLTVIRELYAFTGLTLPOLQ 300
DB 241 VVNEVCSSHVRVIAEAAHLKPPFLQDRYRLRVYVEDLARDPLTVIRELYAFTGLTLPOLQ 300
QY 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
DB 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
QY 361 VVSELEQDLSLDLLPRGMDSPFKWASSTKQPES 395
DB 361 VVSELEQDLSLDLLPRGMDSPFKWASSTKQPES 395

RESULT 2
Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3;
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase G8T-4beta (Corneal N-
DE acetylglucosamine-6-O-sulfotransferase) (Carbohydrate (N-
DE Name=G8T4beta; Synonyms=CHST6;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.175;
RX Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RL family."
RL Glycobiology 11:75-87(2001).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RX Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene."
RL Nat. Genet. 26:237-241(2000).
RN [3]
SEQUENCE FROM N.A.
RP TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Dratchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAG48244.1; -
DR EMBL; AF219991; AAG26327.1; -
DR EMBL; BC074883; AAH74883.1; -
DR EMBL; BC074834; AAH74834.1; -
DR EMBL; AF219990; AAG26325.1; -
DR Genew; HGNC:6938; CHST6.
DR GO; GO:0005794; C:Golgi apparatus; TAS.
DR GO; GO:0005157; P:N-acetylglucosamine 6-O-sulfotransferase ac. .; TAS.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; IC.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 83.0%; Score 1708; DB 2; Length 395;
Best Local Similarity 83.0%; Pred. No. 1.6e-138;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPRFSSTWMLSLMVQGTGILVLSRVQSPSPAGLGRVHVLVLSWRSWSSGSSFGQLF 60
DB 1 MWLPRVSVSTAVTALLLAQTFLLFLVSRPCSPSPAGGEARVHVLVLSWRSWSSGSSFGQLF 60

QY 61 SQHPDVFYLMEPAAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
DB 61 NQHPDVFYLMEPAAWHVWDTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNISDL 120

QY 121 FQWVSRALCSPPVCEAFARGNISSEEVCKPKCATPPGLAOEACSSYSHVVLKEYRFFN 180
DB 121 FQWVSRALCSPPVCEAFARGNISSEEVCKPKCATPPGLAOEACSSYSHVVLKEYRFFN 180

QY 181 LQVLYPLLSDDPALNLRIVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLLSDDPALNLRIVHLVRDPRAVLSRREQTAKALARDNGIVLGTNGTWVEADPGLR 240

QY 241 VVNEVCSSHVRVIAEAAHLKPPFLQDRYRLRVYVEDLARDPLTVIRELYAFTGLTLPOLQ 300
DB 241 VVNEVCSSHVRVIAEAAHLKPPFLQDRYRLRVYVEDLARDPLTVIRELYAFTGLTLPOLQ 300

QY 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
DB 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360

QY 361 VVSELEQDLSLDLLPRGMDSPFKWASSTKQPES 395
DB 361 VVSELEQDLSLDLLPRGMDSPFKWASSTKQPES 395

RESULT 3
Q7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Names-I-GlcNAC-6-ST;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Intestine;
RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RA Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RT sulfotransferase that is highly restricted to intestinal tissue."
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]

```


RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.,
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1; -;
DR EMBL; AF176838; AAD56000.1; -;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;
Query Match 74.7%; Score 1536.5; DB 2; Length 390;
Best Local Similarity 76.0%; Pred. No. 9.4e-124;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
Qy 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFVQOL 59
Db 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFVQOL 59
Qy 60 FSOHPDVFYLMPEPAHWVWDTLSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISD 119
Db 60 FSOHPDVFYLMPEPAHWVWDTLSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISD 119
Qy 120 LFOVAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 179
Db 120 LFOVAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 179
Qy 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTGWVADPRL 239
Db 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTGWVADPRL 239
Qy 240 RVNEVCRSVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPOL 299
Db 240 RVNEVCRSVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPOL 299
Qy 300 QTWHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGALQLLYR 359
Db 300 QTWHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGALQLLYR 359
Qy 360 SVHSELEQRDLSDLLPRGMDSPKAS 387
Db 360 SVHSELEQRDLSDLLPRGMDSPKAS 387
Qy 361 PVYSADQQRDLTLDVLRPGDPHFSWAS 398
Db 361 PVYSADQQRDLTLDVLRPGDPHFSWAS 398
RESULT 4
ID Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9; Q9UBV3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akana T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dotsa A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.;
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAG38023.1; -;
DR EMBL; AF219991; AAG36326.1; -;
DR Genew; HGNC:1973; CHST5.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;
Query Match 74.7%; Score 1536.5; DB 2; Length 411;
Best Local Similarity 76.0%; Pred. No. 1e-123;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
Qy 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFVQOL 59
Db 22 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVPPSPAGLGERVHVLLVSSWRSGSFVQOL 81
Qy 60 FSOHPDVFYLMPEPAHWVWDTLSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISD 119
Db 82 FSOHPDVFYLMPEPAHWVWDTLSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISD 141
Qy 120 LFOVAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 179
Db 142 FFWATSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVWLKEVRFF 201
Qy 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTGWVADPRL 239
Db 202 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTGWVADPRL 261
Qy 240 RVNEVCRSVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPOL 299
Db 262 RLREVCRSVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPOL 321
Qy 300 QTWHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGALQLLYR 359
Db 322 EAWHNTHTGSGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRVQEVCGALQLLYR 381
Qy 360 SVHSELEQRDLSDLLPRGMDSPKAS 387
Db 382 PVYSADQQRDLTLDVLRPGDPHFSWAS 409
RESULT 5
ID Q6GN39 PRELIMINARY; PRT; 392 AA.
AC Q6GN39;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE MGC83048 protein.
GN Name=MGC83048;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=2438257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.B., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC073681; AAH73681.1; -;
GO; GO:0008146; P:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 392 AA; 45969 MW; 7712810F8B4704D7 CRC64;

Query Match 57.7%; Score 1187; DB 2; Length 392;
Best Local Similarity 56.6%; Pred. No. 1.2e-93;
Matches 219; Conservative 77; Mismatches 92; Indels 4; Gaps 3;

QY 8 STVMSLLMVQTGILVLSRQ--VPSAPAGLGERVHVLSSWRSGSFVGLFSQHPD 65
DB NVTAGFLLQITFLIILYSRHTVLPDTEKT-EKVHLLLSWRSGSFVGLFSQHPD 66
QY 66 VFVLMPEAHVMDTSLSGSPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNI 125
DB VFVLMPEAHVMSMFQNNKVLHMAVRDLIRSVFLCDMSVFDAYIPNKNVSELFQWAV 126
QY 126 SRALCSPPVCEAFAGNISSEVCKPLCATRPGLAQACSSVSHVLFQVLS 195
DB SRALCSVPACSHDFREAITNCTVCKIKNPKSIEESNTSHVILKEVRFDFLKVLY 186
QY 186 PLLSDPALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWEADPRLRVNEV 245
DB PLLTDPSLNKKIHLVRDPAVAKRSQAMKYLTRDNGIVLNTNGTKID-DIRYVWREI 245
QY 246 CRSHVRIAEALHKPPFPLODRYLRVYEDLARDPLTVIRELYAFTGLTPOLQW 305
DB CQSHAQMYETAMDKAPSLKGRYMLVRYEDVVRDPLREINQMYEFANLKTAKLKN 305
QY 306 ITHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQLGYSV 365
DB ITHGVGPGTKKEBFQTTSRNAVNSQAWRKDLGFKQIKQIQTICKSEMNLLGYQF 365
QY 366 EQRLSLDLLLPRGMDSEFKWASSTEQ 392
DB ERKDSMDFVLPKRYQFSLPNKEKK 392

RESULT 6
ID QBNCGS PRELIMINARY; PRT; 386 AA.
AC QBNCGS;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Hypothetical protein FLJ90265.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,

RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074746; BAC11177.1; -;
DR GenBank; HGNC:1972; CHST4.
DR GO; GO:0008146; P:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 49.7%; Score 1022.5; DB 2; Length 386;
Best Local Similarity 56.7%; Pred. No. 1.8e-79;
Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;

QY 1 MRLPRESSTVMSLLMVQTGILV--LVSROVPS-SPAGLGERVHVLSSWRSGSFV 56
DB 1 MLLPK--KMKLLLFVLSQMAILALFFHMYSHNLSLSKMAQPERMHVLSWRSGSFV 58
QY 57 GOLFSQHPDVFYLMPEAHVMDTSLSGSPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
DB GOLFGQHPDVFYLMPEAHVMDTSLSGSPALHMAVRDLIRSVFLCDMSVFDAYMEPGR 118
QY 116 NISDLFQWAVSRALCSPPVCEAFAGNISSEVCKPLCATRPGLAQACSSVSHVVLKE 175
DB 119 ROSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVKEACRSYSHVVLKE 178
QY 176 VRFNQLQVLYPLSLDPAALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWV-E 234
DB 179 VRFNQLQVLYPLSLDPAALNLRVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWV-E 238
QY 235 ADPLRVNNEVCRSHVRIAEALHKPPFPLODRYLRVYEDLARDPLTVIRELYAFTGLG 294
DB EQPYYVMQVICOQSLQEIYK-TIQLPKALQERYLVRVEDLARADVAQTSRMYEPVGL 297
QY 295 LTPOLQWTHNTHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
DB 298 FLPHLQWTHNTHGSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRVQELCGGALQ 355
QY 355 LLGYSRVHSELEQRLSLDL 375
DB 356 LLGYSRVHSELEQRLSLDL 376

RESULT 7
ID QY5R3 PRELIMINARY; PRT; 386 AA.
AC QY5R3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
DE sulfotransferase GST-3).
GN Name=GST3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Tonsil;

RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.75;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.;
 RT "Chromosomal localization and genomic organization of the galactose/N-
 RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
 RL family";
 RN Glycobiology 11:75-87(2001).
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2132592; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
 RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellices L.G.,
 RA Rabuka D., Hindsgaul O., March J.D., Lowe J.B., Fakuda M.;
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RT Core1 extension beta 1,3-N-acetylglucosaminyltransferase";
 RL Cell 105:957-969(2001).
 DR EMBL; AF131235; AA033015.1; -;
 DR EMBL; AF280088; AAG48246.1; -;
 DR EMBL; AF149793; AAK48417.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase_1; 1.
 KW Selectin; Transferase.
 SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 49.5%; Score 1017.5; DB 2; Length 386;
 Best Local Similarity 56.4%; Pred. No. 4.7e-79;
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSTVMSLLMVTGTLVF---LVSRQVPS-SPAGLGERVHVLLVSSWRSGSFV 56
 Db 1 MLLPK--KMKLLFLVSQMAILALFFHMYSHNISLSMKAPQRMHVLLVSSWRSGSFV 58
 Qy 57 GOLFSQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFSQHPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLRSVFLCDMDVFDAYL-PWRR 118
 Qy 116 NISDLFQAVSRALCSPVPCAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKE 175
 Db 119 QSSSLFQWNSRALCSAPACDIIPQDEIIIPRAHCRLLCSQPFVVEKACRSYSHVVLKE 178
 Qy 176 VRFNQLVYPLSDPALNLRIHVLRDPRAVLSRQVATKALARDNGIVLGTGTWV-E 234
 Db 179 VRFNQLVYPLSDPALNLRIHVLRDPRAVLSRQVATKALARDNGIVLGTGTWV-E 238
 Qy 235 ADPLRVNVECRSHVRIAEEALHKKPPFPLODRYLVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPYVMQVTCOSQLEIYK-TIQLPKALQERYLLVRYEDLARDPLTVIRELYAFTGLG 297
 Qy 295 LTPOLQWTHNITGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQ 354
 Db 298 FLPHLQWTHNITGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQ 355
 Qy 355 LLGYRSHVSELEQRLDLSLDLL 375
 Db 356 LLGYRSHVSELEQRLDLSLDLL 376

RESULT 8
 Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46;
 DT 01-MAR-2003 (TRENBLrel. 23, Created)
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE CHST4 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max A.I., Wang J., Haieff F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase_1; 1.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 49.1%; Score 1010.5; DB 2; Length 370;
 Best Local Similarity 60.2%; Pred. No. 1.8e-78;
 Matches 204; Conservative 38; Mismatches 92; Indels 5; Gaps 4;

Qy 39 ERHVLLVLSWRSGSFVQGLFSQHDPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLRS 98
 Db 25 ERHVLLVLSWRSGSFVQGLFSQHDPDVFYLMPEAWHVDLTSQGSAPALHMAVRDLRS 84
 Qy 99 VFLCDMDVFDAYL-PWRRNISDLFQAVSRALCSPVPCAFARGNISSEVCKPLCATR 157
 Db 85 VFLCDMDVFDAYL-PWRRNISDLFQAVSRALCSPVPCAFARGNISSEVCKPLCATR 144
 Qy 158 FGLAQACSSYSHVVLKEVRFNQLVYPLSDPALNLRIHVLRDPRAVLSRQVATK 217
 Db 145 FEVVEKACRSYSHVVLKEVRFNQLVYPLSDPALNLRIHVLRDPRAVLSRQVATK 204
 Qy 218 LARDNGIVLGTGTWV-EADPLRVNVECRSHVRIAEEALHKKPPFPLODRYLVRYEDL 276
 Db 205 LMDSIRVMQVTCOSQLEIYK-TIQLPKALQERYLLVRYEDL 263
 Qy 277 ARPLTVIRELYAFTGLTPOLOTWTHNITGSGPGARREAFKTTSRDALSVSQAWRHT 336
 Db 264 ARPVAQTSRMTEFVGLGFLPHLQWTHNITGSGPGARREAFKTTSRDALSVSQAWRHT 321
 Qy 337 LPFAKIRRVQELCGGALQQLGYRSHVSELEQRLDLSLDLL 375
 Db 322 LPYKVSRLQKACGDANMLLYGRHVRSELEQRLDLSLDLL 360

RESULT 9
 Q9RI11 PRELIMINARY; PRT; 388 AA.
 AC Q9RI11;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE L-selectin ligand sulfotransferase (Chst4 protein).

GN Name=Chst4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99361934; PubMed=10435581; DOI=10.1016/S1074-7613(00)80083-7;
RA Hiraoka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
RL Immunity 11:79-89(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Haile S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109155; AAD45579.1; -;
DR EMBL; BC057886; AAH57886.1; -;
DR MCD; MGI:1349479; Chst4.
DR GO; GO:0016021; C: integral to membrane; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Selectin; Transferase.
SQ SEQUENCE 388 AA; 44635 MW; 6D5371AFB6884AEE CRC64;

Query Match 48.2%; Score 991; DB 2; Length 388;
Best Local Similarity 55.2%; Pred. No. 9.1e-77;
Matches 208; Conservative 43; Mismatches 114; Indels 12; Gaps 6;

QY 2 RLPRF-SSTVMSLILMVTOTGILVFLVSQVPSPAGLGERVHVLVLSWSGSGFVQOLF 60
DB 8 RLMLFLGSGQIVVALFTHMSVHRHLSQRESRP-----VHVLVLSWSGSGFVQOLF 61
QY 61 SQHPDVFYLMPEAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRRNISD 119
DB 62 GQHPDVFYLMPEAWHVWMTTSSSTAWKLHMAVRDLIRSVFLCDMSVFDAYMFGPKQSS 121
QY 120 LFQNAVSRALCSPVPVCAFARGNISSEVCKPLCATPFGFLAQACSSYSHVVLKEVRFF 179
DB 122 LFQWEQSRALCAPVCDDFFPAHEITSSPKHCKLLCGQQPFQFMVEKACKSHGFVVLKEVRFL 181
QY 180 NLQVLYPLLDPAINLRIHVLVDRPRAVLSREOTAKALARDNGIVLGTN-GTWVADPR 238
DB 182 SLQALYPLLDPSLNLHVLHVRDPRVAFVRSREHTTILVVDHSHVIGQHLETKEDQP 241
QY 239 LRVVNEVCRSHVRIAEAAHKPPFPFLQDRYRLVRYEDLARDPLTVIRELYAFGLGLTPQ 298

DB 242 YYAMKIICKSQDIVK-AIQTLPEALQQRYFLRYEDLVRAPIAQTRLYKFKVGLDFLPH 300
QY 299 LQTVIHNIHSGPGARREAFKTTSDALSVSQAWRHITLPEAKIRRVQELCGGALQLLGY 358
DB 301 LQTVHNVTRGKGMG--QHAFTNARNALNVSQAWRWSLPYEKVSQLQDQACGAMDLLGY 358
QY 359 RSVHSELEQRDLSLDLL 375
DB 359 LQVRSQOEQGNLSLDLL 375
RESULT 10
Q9WUE5 PRELIMINARY; PRT; 388 AA.
ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (Mus musculus adult male
DE tongue cDNA, RIKEN full-length enriched library, clone:2310003G18
DE product:carbohydrate (chondroitin 6/keratan) sulfotransferase 4, full
DE insert sequence).
DE Name=Chst4;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
 RT "RIKEN integrated sequence analysis (RISA) system-384-formac
 sequencing pipeline with 384 multicapillary sequencer.",
 RL Genome Res. 10:1757-1771(2000).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito K., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF131236; AAD33016.1; -
 DR EMBL; AK009113; BAB26078.1; -
 DR MGD; MGI:1349479; Chst4.
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 388 AA; 44694 MW; DIE9D7796DF8574D CRC64;

Query Match 47.8%; Score 984; DB 2; Length 388;
 Best Local Similarity 54.9%; Pred. No. 3.7e-76;
 Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

QY 2 RLPRF-SSTVMSLLMVTGILVFLVSRQVPSPPAGLGERVHVLVLSWSRSGSFVGLF 60
 DB 8 RLLMFLGQSVIVLAFHMSVHRHLSQREESRRP-----VHVLVLSWSRSGSFVGLF 61
 QY 61 SQHPDVFYLMPEPAHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRRNISD 119
 DB 62 GQHPDVFYLMPEPAHVMWMTFTSTAWKLHMAVRDLIRSVFLCDMSVFDAYNPNPGRQSS 121
 QY 120 LFQAVSRALCPVCEAFAGNTSSSEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 179
 DB 122 LFQEQSRALCSAPVCDFFPAHEISSPKHKLCCQGFDFDVEKACRSHGHVFLKEVRF 181
 QY 180 NLQVLPDLPALNLRIVLVRDPAVLRSEQTAKALARDNGIVLGTN-GTWVEADPR 238
 DB 182 SLQALYPLLTDPNLNHHVHLVRDPAVFRSREHTTIELMVDHSHVLGQHLETIKEEDQP 241
 QY 239 LRVNVECRSHVRIAEALHKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTTPQ 298
 DB 242 YYAMKICKSQVDIVK-AIQLPEALQOYFLRYEDLVRLAPLAQTTLRYKLVGLDFLPH 300
 QY 299 LOTWIHNTHSGGCGARAEAKTTSRDALSQAWRHTLPFAKIRRVQELCGGALLGY 358
 DB 301 LQTVYVYVYVTRGKMG--QHAFTNARNALNVSQAWRSLPYEKVSQLODACEAMDLLGY 358
 QY 359 RSVHSELEQDLSLDLL 375
 DB 359 LQVRSQEQGNLSLDLL 375

RESULT 11
 Q794G9 PRELIMINARY; PRT; 483 AA.
 ID Q794G9
 AC Q794G9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-

DE Sulfotransferase-1).
 GN Name=Gn6st-1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Whole embryos;
 RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
 RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.,
 RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
 sulfotransferase.",
 RL J. Biol. Chem. 273:22577-22583(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Uchimura K., Kadomatsu K., El-Pasakhany F.M., Singer M.S., Izawa M.,
 RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.,
 RT "N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-
 Selectin Ligands and Lymphocyte Homing.",
 RL J. Biol. Chem. 279:35001-35008(2004).
 DR EMBL; AB011452; BAA32139.1; -
 DR EMBL; AB125058; BAA32137.1; -
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR GO; GO:0008146; F:sulfotransferase activity; IDA.
 DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FEFD70E CRC64;

Query Match 31.1%; Score 640; DB 2; Length 483;
 Best Local Similarity 36.2%; Pred. No. 2.1e-45;
 Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERVHVLVLSWSRSGSFVGLFQHPDVFYLM 71
 DB 89 VPAAGSAGAAASLGNATRGTRGGDKQLVYVFTTWSRSGSPFGLFNQNPVFFLYE 148
 QY 72 PAHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-P---WRNISDL--FQWAVS 126
 DB 149 FVHVWQKLYFGDAVSLQGAARDMLSAFYRCDLSVFLYSPAGSGGRNLTTLGTFGAATN 208
 QY 127 RALCSPVCEAFARG--NISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFNQLVL 184
 DB 209 KVCSSPLCPAYRKEVVGLVDDVCKK-CPQRLARFEECKRYTVVKGVRVFDVAVL 267
 QY 185 YPLLSDDPALNLRIVLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPR----- 238
 DB 268 APLLKDPALDLKVHLVRDPAVASSRIRSRHGLIRESLQVRSR-----DPAHRMPF 321
 QY 239 -----LRVNEVCRSHVRIAEALHKPPFLQDRYRLVRYED 275
 DB 322 LEAAGHKLGAKEGMPADYHALGMEVTCNSNAKTLQATAL-QPPDWLQGHYLVRYED 380
 QY 276 LARDPLTVIRELYAFTGLTLPQLOTWIHNTHSGGCGARAEAKTTSRDALSQAWRHT 335
 DB 381 LVGDPVKTLLRVYDFVGLLVSPMEQFALNMTSGSSSK--PPVARSATQANAWRT 438
 QY 336 TLPAKIRRVQELCGGALLGYRSHVHSELEQDLSLDLL 375
 DB 439 ALTFQQLKQVEEFCYQPMVLYGVVNSPEEVKDSKTL 478

RESULT 12
 O88276 PRELIMINARY; PRT; 530 AA.
 ID O88276
 AC O88276;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Names=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=9830482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannegi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR ENBL; AB011452; BAA32138.1; -;
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C3638C CRC64;

Query Match 31.1%; Score 640; DB 2; Length 530;
Best Local Similarity 36.2%; Pred. No. 2.4e-46;
Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERHVLVLSWSRSGSFVGLQSFQHPDVFYLM 71
Db 136 VPAAGSAGAAASLGNAATRGGGDKQLVYVFTTWSRSGSFVGLQSFQHPDVFYLM 195
QY 72 PAHVVDTLTSGSAPALHMAVRLIRSVLCMDVFDAYLP---WRNISDL--FQWAVS 126
Db 196 PVHVMQKLYPGDAVSLQGAARDMLSAFYCDLSVFLQSPAGSGGRNLTTLGFGAATN 255
QY 127 RALCSPPVCEAFARG--NISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFNLQVL 184
Db 256 KVCSSPLCPAYRKEVGLVDDRVCKK-CPQRLARFECECRKYRTLVKGVFVAVL 314
QY 185 YLLSDPALNLRVHVRDPAVLRSEOTAKALARDNGIVLTNGTWTWEADPR----- 238
Db 315 APLKQALDQKVIHLVRDPAVLRSEOTAKALARDNGIVLTNGTWTWEADPR----- 368
QY 239 -----LRVNEVCVSHVIAEALHKKPPFLQDRYLYRYED 275
Db 369 LEAAGHKLGAKEGMPGADYHALGMEVICNSMAKTLOTAL-QPPDWLQGHVLYRYED 427
QY 276 LARDPLTVIRELYAFTGLTPOLOTWIHNTGSGPGARREAFKTTSRDALSVQAWRH 335
Db 428 LVGDPVKTLLRVYDFVGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATQAANWRT 485
QY 336 TLFPFAKIRVQELCGGALQLLGYRSHVSELEQDLSLDLL 375
Db 486 ALTFQIKQVEEFCYQPMVGLGVNVSPEEVKDLSTLL 525

RESULT 13
Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Names=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Morley D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; BC051963; AAH51963.1; -;
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 31.0%; Score 637; DB 2; Length 530;
Best Local Similarity 36.0%; Pred. No. 4.3e-46;
Matches 144; Conservative 74; Mismatches 118; Indels 64; Gaps 10;

QY 30 VPSSPAG-----LG-----ERHVLVLSWSRSGSFVGLQSFQHPDVFYLM 71
Db 136 VPAAGSAGAAASLGNAATRGGGDKQLVYVFTTWSRSGSFVGLQSFQHPDVFYLM 195
QY 72 PAHVVDTLTSGSAPALHMAVRLIRSVLCMDVFDAYLP---WRNISDL--FQWAVS 126
Db 196 PVHVMQKLYPGDAVSLQGAARDMLSAFYCDLSVFLQSPAGSGGRNLTTLGFGAATN 255
QY 127 RALCSPPVCEAFARG--NISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFNLQVL 184
Db 256 KVCSSPLCPAYRKEVGLVDDRVCKK-CPQRLARFECECRKYRTLVKGVFVAVL 314
QY 185 YLLSDPALNLRVHVRDPAVLRSEOTAKALARDNGIVLTNGTWTWEADPR----- 238
Db 315 APLKQALDQKVIHLVRDPAVLRSEOTAKALARDNGIVLTNGTWTWEADPR----- 368
QY 239 -----LRVNEVCVSHVIAEALHKKPPFLQDRYLYRYED 275
Db 369 LEAAGHKLGAKEGMPGADYHALGMEVICNSMAKTLOTAL-QPPDWLQGHVLYRYED 427
QY 276 LARDPLTVIRELYAFTGLTPOLOTWIHNTGSGPGARREAFKTTSRDALSVQAWRH 335
Db 428 LVGDPVKTLLRVYDFVGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATQAANWRT 485
QY 336 TLFPFAKIRVQELCGGALQLLGYRSHVSELEQDLSLDLL 375
Db 486 ALTFQIKQVEEFCYQPMVGLGVNVSPEEVKDLSTLL 525

RESULT 14
Q9UED5
ID Q9UED5
AC Q9UED5;

PRELIMINARY; PRT; 483 AA.

01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
Name=GN6ST;
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
"Human N-acetylglucosamine-6-O-sulfotransferase involved in the
biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
mapping, and expression in various organs and tumor cells.";
J. Biochem. 124:670-678(1998).
[2]
SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.;
"Functional expression and genomic structure of human N-
acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
acetylglucosamine at the nonreducing end of an N-acetylglucosamine
sequence.";
Biochim. Biophys. Acta 1523:269-276(2000).
RL Biochim.
DR ENBL; AB014679; BAA34265.2; -
DR ENBL; AB021124; BAB16886.1; -
DR ENBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;
Query Match 30.9%; Score 636.5; DB 2; Length 483;
Best Local Similarity 37.1%; Pred. No. 4.2e-46;
Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;

Qy 36 GLGE-RVHVLVLSWRSGLSVGQVFSQHPDVFYLMPEPAWHVMDTLSSQSAVALHMAVRD 94
Db 112 GVGDQRQLVYVFTTWRSGSSFFGELFNQNPVEFVFWVWQKLYPGDAVSLQGAARD 171

Qy 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NISSEE 147
Db 172 MLSALYKCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLVDDR 231

Qy 148 VCKPLCATRPFGLAQEACSSYSHVLYKEVFFNQLVPLLSDPALNLRIVHLVRDPRV 207
Db 232 VCKK-CPQRLARPEEBCRYKRTLVKGVRFVAVLAPLRLDPALDLKVIHLVRDPRV 290

Qy 208 LRSEQTAKALRNGIVLGTNGTWVEADPR----- 238
Db 291 ASSRIRSRHGLIRESLQVVRSR-----DPRHRMPFLEAGHKLGAKGEGVGGPADYHA 344

Qy 239 LRVNVECRSHVRIAEALHKPPFPLODRVRLVRYEDLARDPLTVIRELYAFTGLTTPQ 298
Db 345 LGAMEVICNSMAKTLQTL-QPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLSPE 403

Qy 299 LQTIWNIHTHSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRYOELCGGALQLLY 358
Db 404 MEQFALNMTSGSGSSK--PFVVSARNATQAAANWRTALTQQIKQVEEFCYQPMVGLY 461

Qy 359 RSVHSELEQRDLSLDLL 375
Db 462 ERVNSPEVKDLSKTL 478

RESULT 15
Q9Y4C5 PRELIMINARY; PRT; 530 AA.
AC Q9Y4C5; Q9GZNS; Q9Y6P2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE (Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain;
RC MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
"Human N-acetylglucosamine-6-O-sulfotransferase involved in the
biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
mapping, and expression in various organs and tumor cells.";
J. Biochem. 124:670-678(1998).
[2]
SEQUENCE FROM N.A.
RP TISSUE=Umbilical vein endothelium;
RC MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
"CHST1 and CHST2 sulfotransferases expressed by human vascular
endothelial cells: cDNA cloning, expression, and chromosomal
localization.";
Genomics 55:345-347(1999).
DR ENBL; AB014680; BAA34266.2; -
DR ENBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;
Query Match 30.9%; Score 636.5; DB 2; Length 530;
Best Local Similarity 37.1%; Pred. No. 4.7e-46;
Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;

Qy 36 GLGE-RVHVLVLSWRSGLSVGQVFSQHPDVFYLMPEPAWHVMDTLSSQSAVALHMAVRD 94
Db 159 GVGDQRQLVYVFTTWRSGSSFFGELFNQNPVEFVFWVWQKLYPGDAVSLQGAARD 218

Qy 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NISSEE 147
Db 219 MLSALYKCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLVDDR 278

Qy 148 VCKPLCATRPFGLAQEACSSYSHVLYKEVFFNQLVPLLSDPALNLRIVHLVRDPRV 207
Db 279 VCKK-CPQRLARPEEBCRYKRTLVKGVRFVAVLAPLRLDPALDLKVIHLVRDPRV 337

Qy 208 LRSEQTAKALRNGIVLGTNGTWVEADPR----- 238
Db 338 ASSRIRSRHGLIRESLQVVRSR-----DPRHRMPFLEAGHKLGAKGEGVGGPADYHA 391

Qy 239 LRVNVECRSHVRIAEALHKPPFPLODRVRLVRYEDLARDPLTVIRELYAFTGLTTPQ 298
Db 392 LGAMEVICNSMAKTLQTL-QPPDWLQGHVLYVRYEDLVGDPVKTLRRVYDFVGLLSPE 450

Qy 299 LQTIWNIHTHSGPGARREAFKTTSDALSVSQAWRHTLPFAKIRRYOELCGGALQLLY 358
Db 451 MEQFALNMTSGSGSSK--PFVVSARNATQAAANWRTALTQQIKQVEEFCYQPMVGLY 508

Qy 359 RSVHSELEQRDLSLDLL 375

Db 509 ERVNSPEEVKLSKTL 525
|:| | :||| ||

Search completed: June 23, 2005, 08:48:57
Job time : 46.3472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 42.6885 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPRFSSTVMSLLMVQTG.....LPRGMSDFKWSSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2057	100.0	395	4	AAY72638	Aay72638 Mouse gly
2	2057	100.0	395	5	AAU11275	Aau11275 Murine int
3	2042.5	99.3	418	5	ABB81557	Abb81557 Mouse int
4	1727	84.0	395	5	ABB81555	Abb81555 Consensus
5	1708	83.0	395	4	AAY72640	Aay72640 Human gly
6	1708	83.0	395	5	ABB81554	Abb81554 Human cor
7	1708	83.0	395	5	AAE15438	Aae15438 Human dru
8	1708	83.0	395	7	ADI21086	Adi21086 Novel hum
9	1708	83.0	395	8	ADL61235	Adl61235 Human tyr
10	1536.5	74.7	390	4	AAY72639	Aay72639 Human gly
11	1536.5	74.7	390	5	ABB81556	Abb81556 Human int
12	1436	69.8	418	3	AAB41947	Aab41947 Human ORF
13	1333	64.8	394	7	ADJ70405	Adj70405 Human hea
14	1022.5	49.7	386	4	AAW93309	Aaw93309 Human pol
15	1022.5	49.7	386	8	ADL30784	Adl30784 Human pro
16	1017.5	49.5	386	2	AAAY39918	Aay39918 Human gly
17	1015.5	49.4	380	5	AAU11274	Aau11274 Human L-s
18	986.5	48.0	386	3	AAAY79219	Aay79219 Human tra
19	984	47.8	388	2	AAAY39919	Aay39919 Mouse gly
20	712	34.6	169	5	ABB81558	Abb81558 Human cor
21	677	32.9	169	5	ABB81559	Abb81559 Human int
22	647.5	31.5	484	2	AAAY31657	Aay31657 Human N-a
23	647.5	31.5	531	5	AAU69414	Aau69414 Lung smal
24	647.5	31.5	531	8	ADR14283	Adr14283 Human NF-
25	640	31.1	483	2	AAAY31656	Aay31656 Mouse N-a

ALIGNMENTS

RESULT 1

AAAY72638
ID AAY72638 standard; protein; 395 AA.

XX AAY72638;

AC AAY72638;

XX 02-MAY-2001 (first entry)

DE Mouse glycosyl sulfotransferase-4 (GST-4).

XX Mouse, glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy;

KW selectin binding inhibitor; gene therapy; inflammation;

KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;

KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;

KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;

KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;

KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;

KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

KW chromosome 8E1.

XX Mus musculus.

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02695.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

XX diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 2; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4

XX gene is found on chromosome 8E1. GST is a type 2 membrane protein useful

XX for inhibiting a binding event between a selectin and a selectin ligand,

XX which comprises contacting the selectin with a non-sulphated selectin

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC ligand, GST and a small molecular agent that inhibits the sulphonation
CC activity of GST. GST is also useful in inhibiting a selectin mediated
CC binding event. GST is useful in gene therapy to treat disorders such as
CC acute or chronic inflammation, systemic lupus erythematosus (SLE),
CC rheumatoid arthritis, polyarthritis nodosa, polymyositis,
CC dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis,
CC myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's
CC disease, adrenalitis, hypoparathyroidism, pernicious anaemia.
CC demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis,
CC myocarditis, regional enteritis, adult respiratory distress syndrome,
CC infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial
CC asthma, hypersensitivity, rheumatic fever and tissue rejection during
CC transplantation
XX
SQ Sequence 395 AA;

Query Match 100.0%; Score 2057; DB 4; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.7e-205;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVQOLF 60
DB 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVQOLF 60
QY 61 SQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
DB 61 SQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
QY 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 180
DB 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 180
QY 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
DB 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
QY 241 VVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
DB 241 VVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
QY 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYG 360
DB 361 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYG 360
QY 361 VVSELEQRDLSLDLLPRGMDSPFKWASSTKQEPES 395
DB 361 VVSELEQRDLSLDLLPRGMDSPFKWASSTKQEPES 395

RESULT 2
AAU11275
ID AAU11275 standard; protein; 395 AA.
XX
AC AAU11275;
XX
DT 12-MAR-2002 (first entry)
XX
DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.
XX
KW Mouse; beta1,3Gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer;
KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
XX
OS Mus musculus.
XX
PN WO200185177-A1.
XX
PD 15-NOV-2001.
XX

PF 10-MAY-2001; 2001WO-US015452.
XX
PR 11-MAY-2000; 2000US-00569320.
XX
PA (BURN-) BURNHAM INST.
PI Fukuda M, Yeh J, Hiraoka N;
DR WPI; 2002-075226/10.
DR N-PSDB; AAS16948.
XX
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GLCNAC 6-sulfotransferase.
XX
XX Claim 28; Fig 10; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1,3-N-
CC acetylglucosaminyltransferase (beta1,3Gnt) or an active fragment, where
CC beta1,3Gnt directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1,3Gnt that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3Gnt,
CC and/or a beta1,3Gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3Gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST
XX
SQ Sequence 395 AA;

Query Match 100.0%; Score 2057; DB 5; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.7e-205;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVQOLF 60
DB 1 MRLPRFSSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLLSSWRSGSFVQOLF 60
QY 61 SQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
DB 61 SQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
QY 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 180
DB 121 FQWAVSRALCSPPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRF 180
QY 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
DB 181 LOVLYPLSDPALNLRIVHLVRDPAVLRSEOTAKALARDNGIVLGTNGTWVEADPRLR 240
QY 241 VVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
DB 241 VVNEVCRSHVRIAEALHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
QY 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYG 360
DB 301 TWIHNITHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYG 360
QY 361 VVSELEQRDLSLDLLPRGMDSPFKWASSTKQEPES 395
DB 361 VVSELEQRDLSLDLLPRGMDSPFKWASSTKQEPES 395

RESULT 3
ABB81557
ID ABB81557 standard; protein; 418 AA.
XX
AC ABB81557;
DT
DT 05-SEP-2002 (first entry)
DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Mus musculus.
XX
PN US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
XX Fukuda MN, Akama TO;
PI
PI WPI; 2002-507643/54.
XX
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
XX Example 5; Page 24-25; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratectomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
XX Sequence 418 AA;
SQ

Query Match 99.3%; Score 2042.5; DB 5; Length 418;
Best Local Similarity 99.7%; Pred. No. 6e-204;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
1 MRLPRFSTVMSLLMVTGTLVFLVSRQVPSAPGLGERVHVLVLSWRSGSPVGVQLP 60
25 MRLPRFSTVMSLLMVTGTLVFLVSRQVPSAPGLGERVHVLVLSWRSGSPVGVQLP 84
61 SQHPDVFYLMPEPAHVMVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNITSDL 120
85 SQHPDVFYLMPEPAHVMVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNITSDL 144
121 FQWAVSRALCSPVPVCEAFARGNISSEEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
145 FQWAVSRALCSPVPVCEAFARGNISSEEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 204
181 LQVLYPLLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
205 LQVLYPLLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 264

QY 241 VVNEVCRSHVRIAEAAALHKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPOLO 300
Db 265 VVNEVCRSHVRIAE-ALHKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPOLO 323
QY 301 TWIHNITGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYS 360
Db 324 TWIHNITGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYS 383
QY 361 VHSLEQDRLSLDLLPRGMDSFKWASSTKQPES 395
Db 384 VHSLEQDRLSLDLLPRGMDSFKWASSTKQPES 418
RESULT 4
ABB81555
ID ABB81555 standard; protein; 395 AA.
XX
AC ABB81555;
XX
DT 05-SEP-2002 (first entry)
XX
DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
XX Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 10 /label= Ala, Thr, Val
FT Misc-difference 13 /label= Ala, Val, Ser
FT Misc-difference 20 /label= Phe, Cys, Gly
FT Misc-difference 39 /label= Ala, Asp, Glu
FT Misc-difference 96 /label= Val, Met, Ile
FT Misc-difference 142 /label= Ala, Thr, Asn
FT Misc-difference 147 /label= Ala, Asp, Glu
FT Misc-difference 159 /label= Thr, Ser, Gly
FT Misc-difference 238 /label= Gly, His, Arg
FT Misc-difference 294 /label= Ser, Thr, Gly
FT Misc-difference 371 /label= Ala, Thr, Ser
FT Misc-difference 380 /label= Leu, Pro, Met
FT Misc-difference 382 /label= Gly, His, Ser
FT Misc-difference 384 /label= Thr, Ser, Lys
FT Misc-difference 390 /label= Ala, Glu
FT Misc-difference 391 /label= Ser, Lys
FT Misc-difference 392 /label= His, Gln
FT Misc-difference 394 /label= Arg, Glu
FT Misc-difference 395 /label= Asn, Ser
XX
XX US2002061562-A1.

PD 23-MAY-2002.
 XX
 PF 09-AUG-2001; 2001US-00927602.
 XX
 PR 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.
 XX
 PA (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX
 PI Fukuda MN, Akama TO;
 XX
 DR WPI; 2002-507643/54.
 XX
 PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX
 PS Example 5; Fig 2A-B; 69pp; English.
 XX
 CC The present invention describes human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratinoplasty or keratectomy. The present sequence
 CC represents a consensus N-acetylglucosamine-6-sulfotransferase which is
 CC given in the exemplification of the present invention
 XX
 SQ Sequence 395 AA;
 Query Match 84.0%; Score 1727; DB 5; Length 395;
 Best Local Similarity 84.7%; Pred. No. 5e-171;
 Matches 333; Conservative 16; Mismatches 44; Indels 0; Gaps 0;
 QY 1 MRLPRFSTVMSLLMVOTGILVLSRVQVPPSPAGLGERVHVLVLSWSRSGSFVQOLF 60
 DB 1 MMLPRFSTVVTXLLLAQTLLLLFLVSRPGSPAGGXRVHVLVLSWSRSGSFVQOLF 60
 QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 SQHPDVFYLMPEPAWHVWDTLSQGSAAALHMAVRDLXSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
 DB 121 FQAVSRALCSPACSAFPRGXISSEVCKPLCARQPFGLAREACRSYSHVVLKEVRFN 180
 QY 181 LQVLYPLSDPALNLRIVHLVROPRAVLRSEOTAKALARDNGVLGTNGTWVEADPRLR 240
 DB 181 LQVLYPLSDPALNLRIVHLVROPRAVLRSEOTAKALARDNGVLGTNGTWVEADPXL 240
 QY 241 VVNEVCRSHVRIAEAAALHKPPPEFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
 DB 241 VVREVCSSHVRIAEAAALXKPPPEFLRGYRLVRVEDLAREPLAEIRALYAFGLXLPQLE 300
 QY 301 TWIHNTHGSGPGARRAFKTTSDALUSVQAWRHHTLPFAKIRRVQELCGGALQLLGYS 360
 DB 301 AWIHNTHGSGPGARRAFKTTSSRNALNUSQAWRHHTLPFAKIRRVQELCGALQLLGYP 360
 QY 361 VHSLEORDLSLDLLPRGWDSEFKWASSTEQP 393
 DB 361 VYSEDEQRLXLDLVLPRGXDXFKWASSTXXXP 393
 RESULT 5
 ID AAY72640 standard; protein; 395 AA.
 XX
 AC AAY72640;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta).
 DE
 XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX
 OS Homo sapiens.
 XX
 XX WO200106015-A1.
 PN
 XX 25-JAN-2001.
 PD
 XX 19-JUL-2000; 2000WO-US019741.
 PF
 XX 20-JUL-1999; 99US-0144694P.
 PR
 PR 13-JUN-2000; 2000US-00593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Rosen SD, Lee JK, Hemmerich S;
 PI WPI; 2001-138471/14.
 XX
 DR N-PSDB; AAD02697, AAD02700.
 DR
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications.
 XX
 XX Claim 3; Fig 4B; 128pp; English.
 PS
 XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
 CC membrane protein useful for inhibiting a binding event between a selectin
 CC and a selectin ligand, which comprises contacting the selectin with a non
 CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 SQ Sequence 395 AA;
 Query Match 83.0%; Score 1708; DB 4; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
 QY 1 MRLPRFSTVMSLLMVOTGILVLSRVQVPPSPAGLGERVHVLVLSWSRSGSFVQOLF 60
 DB 1 MMLPRVSSTAVTALLAQITFLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVQOLF 60
 QY 61 SQHPDVFYLMPEPAWHVWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
 DB 61 NQHPDVFYLMPEPAWHVWDTLSQGSAAALHMAVRDLVRSVFLCDMDVFDAYLPWRRNISDL 120
 QY 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
 DB 121 FQAVSRALCSPACSAFPRGXISSEVCKPLCARQSFGLAREACRSYSHVVLKEVRFN 180

Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWTWEADPRLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWTWEADPCLR 240
Qy 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLO 300
Db 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYGYS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYGYP 360
Qy 361 VHSELEQDLSLDLLPRGMDSPKWSSTKQPS 395
Db 361 VYSEDEQNLALDLVLPRLGNGFTWASSTASHPRN 395
RESULT 6
ABB81554
ID ABB81554 standard; protein; 395 AA.
XX
AC ABB81554;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological; chromosome 16q22.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
PA (FUKU//) FUKUDA M N.
PA (AKAW//) AKAWA T O.
XX
PI Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
DR N-PSDB; ABB89506.
XX
PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
PS Claim 13; Fig 1A-D; 69pp; English.
XX
CC The present sequence represents human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyse sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratotomy
XX
SQ Sequence 395 AA;
Query Match 83.0%; Score 1708; DB 5; Length 395;
Best Local Similarity 83.0%; Pred. No. 4.8e-169;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
Qy 1 MRLPRFSVTMLSLMVOTGILVFLVSRQVPSSPAGLGRVHVLVLSWSRSGSFVGOLF 60

Db 1 MMLPRVSSTAVTALLAQTEFLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVGOLF 60
Qy 61 SQRPDPFYLMEPAWHVMDTILSQSGAPALHMAVRDLIRSVFLCDMDVFDAYLPMWRNISDL 120
Db 61 NQHPDPFYLMEPAWHVMDTILSQSGAAATLHMAVRDLVRSVFLCDMDVFDAYLPMWRNISDL 120
Qy 121 FQAVSALCSPVPCEAFAPAGNISSEBVCCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
Db 121 FQAVSALCSPVPCEAFAPAGNISSEBVCCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWTWEADPRLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWTWEADPCLR 240
Qy 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLO 300
Db 241 VVNEVCRSHVRIAEAAHLKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTPOLE 300
Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYGYS 360
Db 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGGALQLLYGYP 360
Qy 361 VHSELEQDLSLDLLPRGMDSPKWSSTKQPS 395
Db 361 VYSEDEQNLALDLVLPRLGNGFTWASSTASHPRN 395
RESULT 7
AAE15438
ID AAE15438 standard; protein; 395 AA.
XX
AC AAE15438;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human drug metabolising enzyme (DME)-5.
XX
KW Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
KW inflammatory disorder; acquired immune deficiency syndrome; infection;
KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
KW gastrointestinal disorder; metabolic disorder; developmental disorder;
KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW DME-5.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..35 /label= Signal_peptide
FT Peptide 1..32 /label= Signal_peptide
FT Protein 33..395 /note= "Human mature DME-5 protein"
FT Protein 36..395 /note= "Human mature DME-5 protein"
XX
WO200179468-A2.
XX
PD 25-OCT-2001.
XX
PF 12-APR-2001; 2001WO-US011869.
XX
PR 13-APR-2000; 2000US-0197590P.
PR 19-APR-2000; 2000US-0198403P.
PR 28-APR-2000; 2000US-0200185P.
PR 05-MAY-2000; 2000US-020234P.
PR 11-MAY-2000; 2000US-0203509P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX

PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX
 DR WPI; 2002-066363/09.
 DR N-PSDB; AAD24670.
 XX
 XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated with
 PT aberrant expression of DME such as allergy, anemia, asthma, infertility.
 XX
 XX Claim 1a; Page 131-132; 143pp; English.
 XX
 XX The invention relates to human drug metabolising enzymes referred as DME
 CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the
 CC invention are useful for assessing toxicity of test compounds and in gene
 CC therapy. Sequences of the invention are useful in the diagnosis,
 CC prevention and treatment of autoimmune/inflammatory disorders such as
 CC acquired immune deficiency syndrome (AIDS), adult respiratory distress
 CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune
 CC hemolytic anaemia, contact dermatitis, Crohn's disease,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,
 CC rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral,
 CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell
 CC proliferative disorders such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's
 CC syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine
 CC disorders such as disorders of the hypothalamus and pituitary resulting
 CC from lesions such as primary brain tumours, adenomas, infarction
 CC associated with pregnancy, aneurysms, vascular malformations; eye
 CC disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa
 CC ; metabolic disorders such as Addison's disease, cystic fibrosis,
 CC diabetes, goitre, glycogen storage diseases, hypercholesterolaemia,
 CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,
 CC mannoidosis, obesity; gastrointestinal disorders such as dysphagia,
 CC gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein
 XX
 SQ Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 5; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPFSSTVMSLLMVQGTGLVFLVSRQVPSPPAGLGERVHVLVLSWRSGSFGQLF 60
 DB 1 MWLPRVSTATVALLAQTFELLFLVSRQVPSPPAGLGERVHVLVLSWRSGSFGQLF 60

QY 61 SQHPDVFYLMPEAWHVVMTLSSQGSAPALHMAVRDLIRSVFLCDMDVFDALPWRNLSL 120
 DB 61 NQHPDVFYLMPEAWHVVMTLSSQGSATLHMAVRDLVRSVFLCDMDVFDALPWRNLSL 120

QY 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCNTRPFLGAQACSSYSHVLKEVRFN 180
 DB 121 FQAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVLKEVRFN 180

QY 181 LOVLVPLLSPPALNLRVHVRQVPRVLRREQTAKALRDNGVLGTNGTWVEADPGLR 240
 DB 181 LOVLVPLLSPPALNLRVHVRQVPRVLRREQTAKALRDNGVLGTNGTWVEADPGLR 240

QY 241 VVNEVCRSHVRIAEEAALHKPPFPLQDRYRLVRYVEDLARDPLTVIRELYAFTGLTLPQL 300
 DB 241 VVNEVCRSHVRIAEEAATLKPPFPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300

QY 301 TWIHNTHGSGPGARRAFKTTSDALSQVQRHTLPFAKIRRVQBLCGALQLLGYS 360
 DB 301 AWIHNTHGSGPGARRAFKTTSRNALNSQVQRHALPFAKIRRVQBLCGALQLLGYP 360

QY 361 VHSELEQRDLSDLLPRGMDSFKWASSTKOPES 395
 DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 8
 ADI21086
 ID ADI21086 standard; protein; 395 AA.
 XX
 AC ADI21086;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Novel human protein #61.
 XX
 DE forensic; nutritional source; damaged tissue; diseased tissue;
 KW myeloid cell disorder; lymphoid cell disorder;
 KW bone cartilage tissue growth; tendon tissue growth;
 KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;
 KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003025148-A2.
 XX
 PD 27-MAR-2003.
 XX
 PF 19-SEP-2002; 2002WO-US029964.
 XX
 PR 19-SEP-2001; 2001US-0323739P.
 PR 13-SEP-2002; 2002US-00323739.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;
 PI Haley-Vicente D;
 XX
 XX WPI: 2003-354603/33.
 DR N-PSDB; ADI21802.
 XX
 PT New polynucleotides and secreted proteins, useful for treating myeloid or
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
 PT tissue growth or regeneration, in wound healing, and in tissue repair and
 PT replacement.
 XX
 PS Claim 20; SEQ ID NO 337; 156pp; English.
 XX
 CC The invention relates to an isolated polynucleotide encoding a
 CC polypeptide with biological activity. The polynucleotides and
 CC polypeptides are useful in diagnostics, forensics, gene mapping,
 CC identification of mutations responsible for genetic disorders and other
 CC traits, to assess biodiversity, as nutritional sources or supplements.
 CC The polynucleotides may also be used as molecular weight markers,
 CC chromosome markers or map related gene positions, or as an antigen to
 CC raise anti-DNA antibodies or elicit immune response. The polypeptides are
 CC useful for raising antibodies, as markers for tissues in which the
 CC corresponding polypeptide is expressed, for re-engineering damaged or
 CC diseased tissues, for treating myeloid or lymphoid cell disorders, in
 CC bone cartilage, tendon, ligament and/or nerve tissue growth or
 CC regeneration, in wound healing, in tissue repair and replacement, in
 CC healing of burns, incisions and ulcers, and in treating cancer. The
 CC present sequence represents the amino acid sequence of a novel human
 CC protein.
 XX
 SQ Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 7; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MRLPFSSTVMSLLMVQGTGLVFLVSRQVPSPPAGLGERVHVLVLSWRSGSFGQLF 60
 DB 1 MWLPRVSTATVALLAQTFELLFLVSRQVPSPPAGLGERVHVLVLSWRSGSFGQLF 60

Db 1 MWLPRVSVSTAVTALLLAQTFLLLFLVSRPGPSSPAGGEARVHVLVLSWRSVGGSFVGLQF 60
 Qy 61 SQHPDVFYLMPEAWHVDLTLQSGAPALHMAVRDLIRSVFLCDMDVDFDYLPRRNISDL 120
 Db 61 NQHPDVFYLMPEAWHVDLTLQSGAAALHMAVRDLIRSVFLCDMDVDFDYLPRRNISDL 120
 Qy 121 FQAVSRALCSPVPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Db 121 FQAVSRALCSPVPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLTNGTWTWEADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLTNGTWTWEADPGLR 240
 Qy 241 VVNEVCRSHVRIAEAAHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
 Db 241 VVNEVCRSHVRIAEAAHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
 Qy 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYGYS 360
 Db 301 AWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYGYS 360
 Qy 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQPS 395
 Db 361 VYSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 9

ADL61235
 ID ADL61235 standard; protein; 395 AA.

XX AC ADL61235;
 XX DT 03-JUN-2004 (first entry)
 XX DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
 XX KW predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
 KW vasotropic; vulnary; pharmacogenomic; drug sensitivity; breast cancer;
 KW hypervascular disease; angiogenesis; wound healing scar; human;
 KW biomarker; carbohydrate sulphotransferase 6; enzyme.
 XX OS Homo sapiens.
 XX PN WO2004020583-A2.
 XX PD 11-MAR-2004.
 XX PF 26-AUG-2003; 2003WO-US026491.
 XX PR 27-AUG-2002; 2002US-0406385P.
 XX PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PI Huang F, Han X, Reeves KA, Amler L, Fairchild CR, Lee FY;
 PI Shaw P;
 XX DR WPI; 2004-239171/22.
 XX DR N-PSDB; ADL61098.

XX PT New predictor sets with a plurality of polynucleotides and/or
 PT polypeptides whose expression pattern predicts cell response to a
 PT compound that modulates protein tyrosine kinase activity, useful in
 PT treating breast cancer.
 XX PS Claim 9; SEQ ID NO 159; 649pp; English.

XX CC The invention relates to a novel predictor set comprising a plurality of
 CC polynucleotides and/or polypeptides whose expression pattern is
 CC predictive of the response of cells to treatment with a compound that
 CC modulates protein tyrosine kinase activity, or members of the protein
 CC tyrosine kinase pathway. The molecules of the invention demonstrate
 CC cytostatic, antiangiogenic, vasotropic and vulnerary activities and may

CC be useful in the field of pharmacogenomics, in particular for determining
 .CC drug sensitivity and in treating breast cancer, hypervascular diseases,
 CC angiogenesis and scars in wound healing. The current sequence is that of
 CC a human protein tyrosine kinase biomarker protein of the invention.

XX SQ Sequence 395 AA;

Query Match 83.0%; Score 1708; DB 8; Length 395;
 Best Local Similarity 83.0%; Pred. No. 4.8e-169;
 Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MWLPRVSVSTAVTALLLAQTFLLLFLVSRPGPSSPAGGEARVHVLVLSWRSVGGSFVGLQF 60
 Db 1 MWLPRVSVSTAVTALLLAQTFLLLFLVSRPGPSSPAGGEARVHVLVLSWRSVGGSFVGLQF 60
 Qy 61 SQHPDVFYLMPEAWHVDLTLQSGAPALHMAVRDLIRSVFLCDMDVDFDYLPRRNISDL 120
 Db 61 NQHPDVFYLMPEAWHVDLTLQSGAAALHMAVRDLIRSVFLCDMDVDFDYLPRRNISDL 120
 Qy 121 FQAVSRALCSPVPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Db 121 FQAVSRALCSPVPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVFFN 180
 Qy 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLTNGTWTWEADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSEQTAKALARDNGIVLTNGTWTWEADPGLR 240
 Qy 241 VVNEVCRSHVRIAEAAHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
 Db 241 VVNEVCRSHVRIAEAAHKPPFLQDRYLRVRYEDLARDPLTVIRELYAFTGLGLTQQLQ 300
 Qy 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYGYS 360
 Db 301 AWIHNTHGSGPGARREAFKTTSDALSVSQAWRHLPFAKIRRVQELCGALQLLYGYS 360
 Qy 361 VHSLEQRDLSDLLPRGMDSFKWASSTKQPS 395
 Db 361 VYSEDEQRNLALDLVLRGLNGFTWASSTASHPRN 395

RESULT 10

AAV72639
 ID AAV72639 standard; protein; 390 AA.

XX AC AAV72639;
 XX DT 02-MAY-2001 (first entry)
 XX DE Human glycosyl sulphotransferase-4alpha (GST-4alpha).
 XX KW Human; glycosyl sulphotransferase-4alpha; GST-4alpha; immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.
 XX OS Homo sapiens.
 XX PN WO200106015-A1.
 XX PD 25-JAN-2001.
 XX PF 19-JUL-2000; 2000WO-US019741.
 XX PR 20-JUL-1999; 99US-0144694P.
 XX PR 13-JUN-2000; 2000US-00593828.
 XX PA (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02698, AAD02699.
XX
PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications.
XX Claim 3; Fig 1; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
CC membrane protein useful for inhibiting a binding event between a selectin
CC and a selectin ligand, which comprises contacting the selectin with a non
CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
CC the sulphation activity of GST. GST is also useful in inhibiting a
CC selectin mediated binding event. GST is useful in gene therapy to treat
CC disorders such as acute or chronic inflammation, systemic lupus
CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation. Note: The present sequence is also shown in
CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
XX
XX Sequence 390 AA;
Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 3.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
QY 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 59
DB 1 MWLPRFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 60
QY 60 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 119
DB 61 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 120
QY 120 LFQWASRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 179
DB 121 FFWNATSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 239
DB 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 240
QY 240 RVNVECRSHVRIAEAAALHKPPPPFLQDRIYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 241 RLIREVCRSHVRIAEAAALHKPPPPFLQDRIYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 300
QY 300 QTWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLLYR 359
DB 301 EAWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLLYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSPKAS 387
DB 361 PVISADQQRDLSDLLPRGMDSPKAS 388
RESULT 11
ABB81556
ID ABB81556 standard; protein; 390 AA.
XX
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX

DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
PA (FUKU/) FUKUDA M N.
XX (AKAM/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
DR WPI; 2002-507643/54.
XX
PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX
XX Example 5; Fig 2A-B; 69pp; English.
XX
CC The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratotomy. The present sequence
CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
XX Sequence 390 AA;
Query Match 74.7%; Score 1536.5; DB 5; Length 390;
Best Local Similarity 76.0%; Pred. No. 3.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
QY 1 MRLPRFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 59
DB 1 MWLPRFSSTVMSLLMVQTG-ILVFLVSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 60
QY 60 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 119
DB 61 FSQHPDVFYLMPEPAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPMWRNID 120
QY 120 LFQWASRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 179
DB 121 FFWNATSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 239
DB 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSRQVSPSPAGLGERVHVLVLSWRSGSSFGQL 240
QY 240 RVNVECRSHVRIAEAAALHKPPPPFLQDRIYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 241 RLIREVCRSHVRIAEAAALHKPPPPFLQDRIYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 300
QY 300 QTWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLLYR 359
DB 301 EAWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHLPFAKIRRVQELCGALQLLYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSPKAS 387

XX WPI; 2003-845369/78.
XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
XX Claim 1; SEQ ID NO 2211; 180pp; English.
XX
XX This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, neurotropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytotatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
XX Sequence 394 AA;
XX
XX Query Match 64.8%; Score 1333; DB 7; Length 394;
XX Best Local Similarity 78.1%; Pred. No. 7.2e-130;
XX Matches 257; Conservative 19; Mismatches 47; Indels 6; Gaps 1;
XX
QY 1 MRLPRSSVTWMLSLIMVOTGILVPLVSQVPSPPAGLGERVHVVLVSSWSSGSPVGGOLF 60
DB 1 MWLPRVSTAVTALLLAQTFLFLVLSRPGSPAGGEARVHVVLVSSWSSGSPVGGOLF 60
XX
QY 61 SQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
DB 61 NQHPDVFYLMPEAWHVMWDTLSQGSAAALHMAVRDLVRSVFLCDMDVFDAYLPWRRNISDL 120
XX
QY 121 FQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
DB 121 FQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
XX
QY 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLCTNGTWVEADPLR 240
DB 181 LOVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLCTNGTWVEADPLR 240
XX
QY 241 VVNEVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLTPQLQ 300
DB 241 LIREVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLTPQLQ 300
XX
QY 301 TWIHNTHGSG-----PGARRAEKTTTS 323
DB 301 AWIHNTHGSGASQSPKPSIIRLCMRATS 329
XX
RESULT 14
AAM93309
ID AAM93309 standard; protein; 386 AA.
XX
AC AAM93309;
XX
XX 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 2817.
XX Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
OS
XX
XX EPI130094-A2.
XX

PD 05-SEP-2001.
XX
XX 07-JUL-2000; 2000EP-00114089.
XX
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI; 2001-524255/58.
DR N-PSDB; AAK94229.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
XX Claim 8; SEQ ID NO-2817; 1380pp + Sequence Listing; English.
XX
XX The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a polypeptide encoded by a full length
CC human cDNA of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in CD-ROM
CC format directly from EPO
XX
XX Sequence 386 AA;
XX
XX Query Match 49.7%; Score 1022.5; DB 4; Length 386;
XX Best Local Similarity 56.7%; Pred. No. 1.9e-97;
XX Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;
XX
QY 1 MRLPRSSVTWMLSLIMVOTGILVPLVSQVPSPPAGLGERVHVVLVSSWSSGSPV 56
DB 1 MLLPK--KKMLLFLVSQMAILALFFHMYSHNLSLSMKAKAQPMMHVVLVSSWSSGSPV 58
XX
QY 57 GOLFSQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRR 115
DB 59 GQLFGQHPDVFYLMPEAWHVMWDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRR 118
XX
QY 116 NISDLFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
DB 119 RQSSLFQWENSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 178
XX
QY 176 VRFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLCTNGTWV-E 234
DB 179 VRFFNLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLCTNGTWV-E 238
XX
QY 235 ADPLRVVNEVCRSHVRIAEAAALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
DB 239 EDQPYVMQVICSQSEIYK-TIQSLPKALQERLLVRYEDLARDPLTVIRELYAFTGLG 297
XX
QY 295 LTPQLQTIWNIHNTGSGPGARREAFKTSRDALSVQAWRHTLPFAKIRRVQELCGGALQ 354
DB 298 FLPHLQTIWNIHNTGSGPGARREAFKTSRDALSVQAWRHTLPFAKIRRVQELCGGALQ 355
XX
QY 355 LLGYRSHVSELEQORDLSLDLL 375
DB 356 LLGYRSHVSELEQORDLSLDLL 376
XX
RESULT 15
ADL30784
ID ADL30784 standard; protein; 386 AA.
XX

AC ADL30784;
XX 20-MAY-2004 (first entry)
XX Human protein encoded by a full length cDNA clone SeqID 2817.
DE human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method.
XX Homo sapiens.
XX EPI396543-A2.
XX 10-MAR-2004.
XX 07-JUL-2000; 2003EP-00025638.
XX 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX Example 1; SEQ ID NO 2817; 1340pp; English.
XX This invention relates to a novel primers useful for synthesising full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
XX length human protein of the invention.
XX Sequence 386 AA;
Query Match 49.7%; Score 1022.5; DB 8; Length 386;
Best Local Similarity 56.7%; Pred. No. 1.9e-97;
Matches 216; Conservative 42; Mismatches 112; Indels 11; Gaps 7;
Qy 1 MRLPRFSTVMSLLMYQTGLVFP---LVSRQVPS-SPAGLGERVHVHLVLSWSRSGSFV 56
Db 1 MLLPK--KMKLLFLVSMQAILALFFHMYSHNISLSLMAQPERMHVHLVLSWSRSGSFV 58
Qy 57 GOLFSQHPDVPYLMPEAHVWVDTLSCGAPALHMAVRDLIRSVFLCDMDVDFAYL-PWRR 115
Db 59 GOLFGQHPDVPYLMPEAHVWVDTFKQSTAMMLHMAVRDLIRAVFLCDMSVDFAYMEGPR 118
Qy 116 NISDLFWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
Db 119 QSSSLFQWENSALCSPACDIIIPQDEIIIPRAHCELLCSQOPFEVVEKACRSYSHVVLKE 178
Qy 176 VRFNQLVLYPLSDPALNLRIVHLVDRDPRAVLRSGOTAKALARDNGIVLGTNGTWV-E 234
Db 179 VRFNQLSLYPLLDPSLNLHIVHLVDRPRAVFRSRRRTKGDLMIDNRIVMGQHEQKLKK 238
Qy 235 ADPLRVNVECRSHVRAEAAHKPPFPLODRVLRVYEDLARDPLTVIRELYAFTGLG 294
Db 239 EDQPYVNVQVTCQSOLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMTEFVGL 297
Qy 295 LTPOLQTVIHNITHGSGFGARREAPKTTSRDALSVQAWRHHTLPFAKIRRVQELCGGALQ 354

Db 298 FLPHLOTWVHNITRGKMG--DHAFHTNARDALNVSQAWRWSLPYEKVSRQLKACGDAMN 355
Qy 355 LLGYRSVHSELEQORDLSLDLL 375
Db 356 LLGYRHRVSRSEQQRNLLLDLL 376
Search completed: June 23, 2005, 08:43:22
Job time : 45.6885 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.95931 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPFSSTVMSLLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: Pirl1:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	643.5	31.3	484	2 JE0261	N-acetylglucosamin
2	580	28.2	486	2 JC7351	N-acetylglucosamin
3	577	28.1	484	2 JC7350	N-acetylglucosamin
4	495	24.1	458	2 A57397	chondroitin 6-sulf
5	148.5	7.2	307	2 B95934	probable enzyme, C
6	126	6.1	388	2 G70729	hypothetical prote
7	101.5	4.9	262	2 T35999	probable aminoglyc
8	99	4.8	571	2 AG3174	acetylacetate synth
9	98	4.8	856	2 F87316	conserved hypothet
10	95	4.6	547	2 A64878	peptide transport
11	94.5	4.6	892	1 S42228	replication licens
12	93	4.5	547	2 G90862	hypothetical prote
13	93	4.5	547	2 B85756	hypothetical prote
14	93	4.5	559	2 T03412	malate synthase (E
15	92.5	4.5	327	2 T50744	spheroidene monoox
16	92.5	4.5	4077	2 T17484	hypothetical prote
17	91	4.4	562	2 T48413	malate synthase-li
18	91	4.4	925	2 T02811	DNA excision/repai
19	90	4.4	933	1 DEECOG	oxoglutarate dehyd
20	90	4.4	933	2 F85573	hypothetical prote
21	90	4.4	933	2 G90722	oxoglutarate dehyd
22	89.5	4.4	311	2 B40216	flavonol 3'-sulfit
23	89.5	4.4	459	2 G83174	exodeoxyribonuclea
24	88	4.3	668	2 T01685	crpl protein - mai
25	87.5	4.3	342	2 T49687	lipoyltransferase
26	87.5	4.3	678	2 T23341	beta-catenin - Cae
27	87.5	4.3	860	2 C82750	mannosyltransferas
28	87.5	4.3	1158	2 F90854	probable host spec
29	87	4.2	413	2 H86825	hypothetical prote

30	87	4.2	436	2 E90261	hypothetical prote
31	87	4.2	643	2 B72602	hypothetical prote
32	87	4.2	814	2 B75050	secretory protein
33	87	4.2	1118	2 C95385	probable adenylate
34	86.5	4.2	301	2 D87684	transcription regu
35	86.5	4.2	327	2 S49819	crta protein - Rho
36	86.5	4.2	899	2 B38529	nikB protein - Esc
37	86.5	4.2	1092	2 H69071	DNA-directed DNA p
38	86	4.2	652	2 T45499	probable DNA recom
39	86	4.2	950	2 T38777	probable valyl trn
40	86	4.2	1165	2 D85842	probable tail fibe
41	86	4.2	1185	2 F90877	probable host spec
42	86	4.2	3796	2 T18514	lysosomal traffick
43	85.5	4.2	385	2 S72702	8-amino-7-oxononan
44	85.5	4.2	425	2 E83023	3-deoxy-D-manno-oc
45	85.5	4.2	532	2 S33758	flavin-containing

ALIGNMENTS

RESULT 1

JE0261

N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human

C/Species: Homo sapiens (man)

C/Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002

C/Accession: JE0261

R/Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998

A/Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of (

A/Reference number: JE0261; MUID:98391845; PMID:9722682

A/Accession: JE0261

A/Molecule type: mRNA

A/Residues: 1-484 <UCH>

A/Cross-references: DDBJ:AB014679

C/Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenos

C/Superfamily: chondroitin 6-sulfotransferase

C/Keywords: sulfotransferase

Query Match 31.3%; Score 643.5; DB 2; Length 484;
Best Local Similarity 36.6%; Pred. No. 8.3e-48;
Matches 139; Conservative 77; Mismatches 117; Indels 47; Gaps 9;

Qy	33	SPAGLGRVH-VLVLSWRSGSSFFVGOLFQHPDVFYLMPEAWHVDLTLSQGSAPALHMA	91
Db	110	APEGVGDKRHMMYVFTTWRSQSPFGEFNQNPVEFLYEPVHWQKLYPGDAVSLQA	169
Qy	92	VRDLIRSVFLCDMDVFDAYLP---WRRNISDL---FQWAVSRALCSPPVCEAFARG--NIS	144
Db	170	ARDMLSALYRCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVCSPLCPAYRKEVVGLV	229
Qy	145	SEEVCKPLCATRPGLAQACSSYSHVVLKEVRFNQVLYPLLSDPALNRIYHLVRDP	204
Db	230	DDRVCCK-CPQRLARPEECRKYRTLVIKGRVFDVAVLAPLLRDPALDLKVIHLVRDP	288
Qy	205	RAVLRSREQTAKALARDNGIVLGTNGTWVADPR-----	238
Db	289	RAVASSRIRSRHGLIRESLQVRSR-----DPAHRMPPFLAAGHKLGAKKEGVGGPAD	342
Qy	239	---LRVNVVECRSHVRIAEALHKPPFPFLQDRYRLVRYEDLDARPLTVIRELYAFTGLGL	295
Db	343	YHALGMEVICSNAKTLQTLAL-QPPDWLQGHLYLVRYEDLVGEVPKTLRRYDFVGLV	401
Qy	296	TPQLQTVIHNTHSGGPGARREAFKTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQL	355
Db	402	SPEMEQFALNMTSGSGSSK--PFWASARNATQANAWRTALTTFQIQKQVEFCYQPMNV	459
Qy	356	LGYSRVSELEQORDLSLDLL	375
Db	460	LGYSRVNSPEEVKDLSTLL	479

RESULT 2
 JCT351
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
 C:Species: Homo sapiens (man)
 C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
 C:Accession: JCT351
 R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsuoka, T.; Yamakawa, N.; Kuroawa, N.; Biochem. Biophys. Res. Commun. 274, 291-296, 2000
 A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a cDNA from human liver. J. Biol. Chem. 274, 291-296, 2000
 A:Reference number: JCT350; MUID:20374462; PMID:10913333
 A:Accession: JCT351
 A:Molecule type: mRNA
 A:Residues: 1-486 <UCH>
 A:Cross-references: UNIPROT:Q95667; DBJ:AB040711
 C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation site, is involved in the synthesis of L-selectin ligand in the Golgi apparatus. This enzyme is involved in the synthesis of L-selectin ligand in the Golgi apparatus. This enzyme is involved in the synthesis of L-selectin ligand in the Golgi apparatus.
 C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 28.2%; Score 580; DB 2; Length 486;
 Best Local Similarity 34.8%; Pred. No. 2.7e-42;
 Matches 139; Conservative 56; Mismatches 137; Indels 68; Gaps 8;

QY 4 PRFSSTVMSLLMVQIGILVLSRVQVPSPPAGLGRVHVLVLSWSSGSSFFVGLFSQH 63
 DB 83 PRFPSNL-----SGAVCEAVSR-----EKQHYVHATWRTGSSFLGELFNQH 124

QY 64 PDVYLMPEAHVWDTLSQSSAPALHMAVRDLIRSVFLCDMDVFDAYLP----- 112
 DB 125 PDVYLYEPNHLWQALPGDABSLQALRDLMLSLFRCDPSVLRLYAPPGDPAARAPDT 184

QY 113 WRNNSDLFOWAVSRALCSPVPC-----EAFAGNISSEEVCKPLCATRPGLAQACSSY 168
 DB 185 ANUTTAALFWRNTKVICSPPLCPGAPRAAEVGLVEDTACERSCPPVIRALEACRKY 244

QY 169 SHVVLKEVRFNQLVPLSDPALNRIHVLVRDPAVLRSRQETAKALARDNGIVL-- 226
 DB 245 PVVVKDRLDLGLVPLLRDPLGLNKKVQLFRDPRAVHNSRLKSRQGLRESIQVLR 304

QY 227 -----GTNGTWVEADPR-----LRVNVNEVCRSHVRIAEALH 258
 DB 305 RQRGRDRFHRVLLAHGVGARGPGOSRALPAAPADFFLTGALEVICAEAWRLDLFARGA-- 362

QY 259 KPPFLQDRVRLVRYEDLARDPLTVIRELYAFTGLTLPQLOTHNIHNSGSPGARREA 318
 DB 363 --PAWLRRRYLRYEDLVQPAQRLLRFGLRDLALDAFALNMTGGAAYGADR-P 419

QY 319 FKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGY 358
 DB 420 FHLSDAREAVHAWRRLSREQVRQVEAACAPAMRLLAY 459

RESULT 3
 JCT350
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
 C:Accession: JCT350
 R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsuoka, T.; Yamakawa, N.; Biochem. Biophys. Res. Commun. 274, 291-296, 2000
 A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a cDNA from human liver. J. Biol. Chem. 274, 291-296, 2000
 A:Reference number: JCT350; MUID:20374462; PMID:10913333
 A:Accession: JCT350
 A:Molecule type: mRNA
 A:Residues: 1-484 <UCH>
 A:Cross-references: UNIPROT:Q95N80; UNIPROT:Q95P78; DBJ:AB040710
 C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation site, is involved in the synthesis of L-selectin ligand in the Golgi apparatus. This enzyme is involved in the synthesis of L-selectin ligand in the Golgi apparatus. This enzyme is involved in the synthesis of L-selectin ligand in the Golgi apparatus.
 C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 28.1%; Score 577; DB 2; Length 484;
 Best Local Similarity 34.5%; Pred. No. 4.8e-42;
 Matches 135; Conservative 57; Mismatches 141; Indels 58; Gaps 8;

QY 31 PSSPAGLG-----ERVHVLVLSWSSGSSFFVGLFSQHDPDVFYLMPEAHVWDTLSQ 82
 DB 82 POSPGNLSAVGAVTQEQKHYYHATWRTGSSFLGELFNQHDPDVFYLMPEAHVWDTLSQ 141

QY 83 GSAPALHMAVRDLIRSVFLCDMDVFDAY-----LPMRNRISD--LFQWAVSRALCS 131
 DB 142 GDAESLQALRDLMLSLFRCDPSVLRLYAPPGDPCGERAPDSANLTTAMLFWRTRTKVCS 201

QY 132 PVPCEAFARG-----NISSEVCKPLCATRPGLAQACSSYSHVVLKEVRFNQLVLYPL 187
 DB 202 PFLCPAARADAVGLVEDKACESTCPPVSLRALEACRKYPVVWIKOVRLLDLGLVLYPL 261

QY 188 LSDPALNRIHVLVRDPAVLRSRQETAKALARDNGIVLGTN-----GTWVEAD 236
 DB 262 LRDPGLNLKVVQLFRDPRAVHNSRLKSRQGLRESIQVLRTRQGRDHFRVLLAHGVDR 321

QY 237 P-----LRVNVNEVCRSHVRIAEALHKKPPPLQDRYRLVRYEDL 277
 DB 322 PGQARALPSAPRADFFLTSALEVICAEAWRLDLFTRGA----PAWLRRRYLRLRYEDLV 377

QY 278 RDPLTVIRELYAFTGLTLPQLOTHNIHNSGSPGARREAFKTTSRDALSVQAWRHLP 337
 DB 378 WQFOAQLRRLRFSGRLTLAALDAFALNMTGSAAYGADR-PFHLSDAREAVHVMRRL 436

QY 338 PFAKIRRVQELCGGALQLLGYRSVHSELEOR 368
 DB 437 SOEQVRQVETACAPAMRLLAYPRSGDERDRK 467

RESULT 4
 AS7397
 chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
 C:Accession: AS7397
 R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi, J.; Biol. Chem. 270, 18575-18580, 1995
 A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase. J. Biol. Chem. 270, 18575-18580, 1995
 A:Reference number: AS7397; MUID:95355490; PMID:7629189
 A:Accession: AS7397
 A:Molecule type: mRNA
 A:Residues: 1-458 <FUK>
 A:Cross-references: UNIPROT:Q92179; GB:D49915; NID:g971262; PIDN:BAA08655.1; PID:g971263
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase

Query Match 24.1%; Score 495; DB 2; Length 458;
 Best Local Similarity 36.1%; Pred. No. 5.8e-35;
 Matches 129; Conservative 59; Mismatches 125; Indels 44; Gaps 13;

QY 40 RVHVLVLSWSSGSSFFVGLFSQHDPDVFYLMPEAHVWDTLS--OGSAPALHMAV--RDL 95
 DB 112 RRHVLMMATRTGSSFFVGEFFNQGNIFYLFEPLHIERVTTFEPGGAAGVSAVYRDV 171

QY 96 IRSVFLCDMDVFDAYL---PWRNRISDLFQWAVSRALCSPVPC-----EAFAGNISSE 147
 DB 172 LQQLLLCDLYILESIFSPAPEHLLTAALFRGSSHSLCEEPVCTPSLKKVFEKHCNRR 231

QY 148 VCKPLCATRPGLAQACSSYSHVVLKEVRFNQLVPLSDPALNRIHVLVRDPRPV 207
 DB 232 -CGPLNIT-----LAAEACRKHQKMAKTVRIQLSFLQPLAEDPRDLRLIIQLVRDPRV 286

QY 208 LRSRQETAKALARDNGIVLGTNGTW-----VEADPRLRVNVNEV-----CRSHVRIAEAL 257
 DB 287 LVSRMVAFS-----GKYESNKKWAAEAEAPLQ--EDEVQRLRGNCESIRLSAELGL 335

QY 258 HKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLOTHNIHNSGSPGARRE 317
 DB 336 -RQPRWLGRVYLVRYEDVARAPLRALEMYRFAGIHPTFPQVEEWRANTQAPQDS---N 391

QY 318 AFKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGYRSVHSELEOR 372

Qy 285 RELYAFGLGLTPQ-LOTWIHNITHGSGPGARREAFKTTSRDALSYSQAWRHLPFAKIR 343

Db 103 --ITPSRGVGVVETRTW-----PGARRSAHPQTSFAALGG-----R 138

Qy 344 RVQELCGGALQL-LGYRSVHSELEQRLSLDLLLPRGMD5 382

Db 139 AAEVWAGHATDCRIGERSPLAALERLDARV-LLLGAGYDA 177

RESULT 8

AG3174

acetolactate synthase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid

C;Species: Agrobacterium tumefaciens

C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 16-Aug-2004

C;Accession: AG3174

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutuyavin, T.; Levy, R.; Li, M.; McClellan

i Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A;Reference number: AB2577; MUID:21608550; PMID:11743193

A;Accession: AG3174

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-571 <KUR>

A;Cross-references: UNIPROT:O8UKH9; GB:AE008687; PIDN:AAU45813.1; PID:gl7743551; GSPDB:G

A;Experimental source: strain C58 (Dupont)

C;Genetics:

A;Gene: atrC

A;Genome: plasmid

C;Superfamily: Acetolactate synthase, large subunit/pyruvate oxidase; thiamin pyrophosph

Query Match 4.8%; Score 99; DB 2; Length 571;

Best Local Similarity 19.5%; Pred. No. 1.7;

Matches 70; Conservative 47; Mismatches 114; Indels 128; Gaps 14;

Qy 81 SOGSA-PALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDLFOWAVSR----- 127

Db 151 SSGRAGFAVILLPADLRSV-----STTSRRRTNNLGHWPLDRSPADKDLIDIA 201

Qy 128 -ALCSPPVCEAFARGNISSEVECKPLCATRPFGLAQEACS-----SYSHVLKVEVRFN 180

Db 202 KATANARAPVVIAGGIGHSG-----AAQRLQLQDECIPVFTTMMKGSGVDWHPLS 255

Qy 181 LOVLYPLSDPALNRLVHLVRDPRVLRSEQTAKALADNGIVLGT-----NGT--WVE 234

Db 256 GGVLASLVGPKSLGRHTVEIVREADLV-----VLIGTRNQNGTDNRQ 299

Qy 235 ADPLRLRVNEVCRSHVRIAEAAHKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLG 294

Db 300 YSDAQIIH-----IDTDQEVGRNFPEARLVDARETLAGURERIRLCDLH 346

Qy 295 LTPQLQVTHIHNTHGSGFGARREAFKTTSDAL--SVSQAWR-----HTLPFA 340

Db 347 L-----RTVSREAVCNLRLESWRKFNDDRRGCVYSSEASPLR 382

Qy 341 KIRRVQELG-----GALQLLGYSVHSELEQRLSLDLLLPRGMD5FKW 385

Db 383 PERIMAELOGVIDENTTVVADASYSSMWVLGQLRISSESTR-----VLTFRGLAGLW 435

RESULT 9

F87316

conserved hypothetical protein CC0543 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C;Accession: F87316

R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Frazer, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: F87316

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-856 <STO>

A;Cross-references: UNIPROT:Q9AAQ3; GB:AE005673; NID:gl3421732; PIDN:AAK22530.1; GSPDB:G

C;Genetics:

A;Gene: CC0543

Query Match 4.8%; Score 98; DB 2; Length 856;

Best Local Similarity 19.5%; Pred. No. 3.5;

Matches 78; Conservative 53; Mismatches 152; Indels 118; Gaps 19;

Qy 39 ERVHVLVLSWSRSGS-----SFYQLFSQHPDVFYLMPEAWHV--W-----DTLS 81

Db 470 EMIH-----AWDKGARNVIANVGDIKPAEIGTSHFLEMAWDIDRWKSKQKFLSDWTA 524

Qy 82 QGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDLFOWAVSRALCSPPVCEAPARG 141

Db 525 RNLLGPALAAKATADLLDRYRLNFERPEHLEW-----PPVAENRHLS 566

Qy 142 NISSEVECKPLCATRPFGLAQEACSSYSHVLKVEVRFNQLVLYPLSDPALNRLVHLV 201

Db 567 SYTPKEVSARLRAFRTLVAETKATSOQVPTLQDA-WFEL-VEFPIRISAAANMRFFAAE 624

Qy 202 RDPRAVLRSEQTAKALADNGIVLGTNGTWVEADPRLRVVNEVCRSHVRIAEAAHLH--K 259

Db 625 R-YNALIDGRQ-----AMARSAG-----GAAVEAQAEITALTD--RFHQIAGGKWRWF 671

Qy 260 PPPFLQDRYRLVRYED-----LARDPLTVIRELYAFTGLGLTLPQLQVTHIHNTHGSGP 312

Db 672 PEEPADSQWRIYRARPIPLPGAALTADPAFLAE-----VDGTLFAGSP 715

Qy 313 GARREAFKTTSDALSVSQAWHTLPPAKIRR-----VQELCGGALQL----- 355

Db 716 VFEAEAFK-----ANRGWRFV---EGVGRGDGVMIAAAGADLTLEVEVKTEGRSLR 764

Qy 356 LGYRSVHSELEQRLSLDLLL-----PRGMD5FKWA 386

Db 765 IGVLPFPFGQGGETALDVSINGEAPQVRSWPRAVGSPAWA 805

RESULT 10

A64878

peptide transport periplasmic protein sapA precursor - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C;Accession: A64878

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: A64878

A;Status: nucleic acid

A;Molecule type: DNA

A;Residues: 1-547 <BLAT>

A;Cross-references: UNIPROT:Q47622; GB:AE000227; GB:U00096; NID:gl787543; PIDN:AAC74376.J

A;Experimental source: strain K-12, substrain MGL655

C;Genetics:

A;Gene: sapA

C;Superfamily: dipeptide transport protein

C;Keywords: periplasmic space; transport system

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-547/Product: peptide transport periplasmic protein sapA #status predicted <MAT>

Query Match 4.6%; Score 95; DB 2; Length 547;

Best Local Similarity 22.0%; Pred. No. 3.5;

Matches 109; Conservative 54; Mismatches 170; Indels 162; Gaps 27;

Qy 11 MSLSLMVQGTIL-----VFLVSRQVPS---SPAGLGRVHVVLV 46

Db 4 VLSSLLVIAGLVSGOIAAPSPHAPDIRDSGFVVCVSGVNTFNPFSKASSGLIVDTL-- 61

Db 384 TRSQMWPS-PL-----KTAELIQDMAQGVKVVIVPVEGRFOEARLMDMSHDLTLS--- 435
Qy 381 DSFKWASSTKQPES 395
Db 436 ---GWATDS-NDPDS 446

RESULT 13
885756
hypothetical protein sapA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: B85756
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85756
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-547 <STO>
A:Cross-references: UNIPROT:O8X7F3; GB:A8005174; NID:gl21515483; PIDN:AAG56510.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: sapA
C:Superfamily: dipeptide transport protein

Query Match 4.5%; Score 93; DB 2; Length 547;
Best Local Similarity 22.0%; Pred. No. 5.3;
Matches 109; Conservative 53; Mismatches 171; Indels 162; Gaps 27;
Qy 11 MSLSLMVQTGIL-----VFVLSRVQPS-----SPAGLGERVHVLV 46
Db 4 VLSSLLVIAGLVGQAIAAPESPPIADIRDSGFYVCVSGQVNTFNPSSKSSGLIVDTL-- 61
Qy 47 SSWRSRGSFVGLQPSQHPDVFYLMPEAWHVDTLSCGSAPALHMAVRDL----- 95
Db 62 -----AAQFYRLDLVDVPIYRLMPELAESWEVLNDAATYRFLR-RDVPFQKTAFTPT 115
Qy 96 ----IRSVFLCDMDVFDAYLFWRRNIS-----DLFQWAVSRALCSPVCEAFARGNIS 144
Db 116 RKNMADDVFTFORIFDRNNPW-HNVNGSNFPYFDSLQFA-----DNVK 158
Qy 145 SEEVCKPLCATRPGLEACSSYSHVVLKEVRFNLOVLYPLSDPALNLRIVHLVRDP 204
Db 159 S--VRKLDNHTVEPLAQDPAFLHLATHYASVMSAEYARKLEKED-----ROEQLDROP 212
Qy 205 -----RAVLRREQ-----TAKALARDNGIVLGTNGT-----W 232
Db 213 VGTGPYQLSEVRAGQFIRLQRHDDFWRGKPLMPQVVVDLSGGTGRLSKLLTGECIDVLAW 272
Qy 233 VEA-----DPLRVVNEVCRSHVRIABAAHLKPPPLQD---RYRLVRYEDLARDPL 281
Db 273 PAASQLSILRDDPRLR---TLRPGMVAFLAFTAKPLNPAVRHALA---LAINNQ 325
Qy 282 TVIRELY---AFTGLGTPLOQTWIHN-----ITHGSGPGARRAFKTTSDALSV----- 329
Db 326 RLMQSIYVGAETAASILPLR-ASWAYDNEAKITE-YNPAKSREQLKLLGLENTLKLWVP 383
Qy 330 --SQAWRHTLPFAKIRRVQELCGGALQLLGVRSVHSELEOR-----DLSLDLLPRGM 380
Db 384 TRSQMWPS-PL-----KTAELIQDMAQGVKVVIVPVEGRFOEARLMDMSHDLTLS--- 435
Qy 381 DSFKWASSTKQPES 395
Db 436 ---GWATDS-NDPDS 446

RESULT 14
T03412
malate synthase (EC 4.1.3.2) - maize

C:Species: Zea mays (maize)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C:Accession: T03412
R:Paek, N.C.
submitted to the EMBL Data Library, September 1994
A:Reference number: Z14939
A:Accession: T03412
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559 <PAE>
A:Cross-references: UNIPROT:P49081; EMBL:L35914; NID:g532624; PIDN:AAB04118.1; PID:g5326;
A:Experimental source: strain TX5855; scutellum
C:Genetics:
A:Note: MS
C:Function: <ACG>
A:Description: catalyzes the aldol condensation of glyoxylate with acetyl-CoA to form mal
A:Pathway: glyoxylate cycle
C:Superfamily: malate synthase
C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 4.5%; Score 93; DB 2; Length 559;
Best Local Similarity 22.5%; Pred. No. 5.4;
Matches 56; Conservative 30; Mismatches 95; Indels 68; Gaps 12;
Qy 159 GLAQEACSSYSHVYLKEVRFNLOVL-----YPLSDPALNLRIVHLVRDPRVLRSE 212
Db 312 GMAQHFMRSYSHLLIHTCHRRGVHANGMAAQIPKIDAAANEAALELYRKDKL----RE 367
Qy 213 QTAKALARDNGIVLGTNGTWEADPLRVVNEVCRSHV-----RIABAAHLKPPPLQDR 267
Db 368 VRA-----GHGDTWAAHPGLIPAIRVFEHGLGRPNQIGDAAGHGAS----- 411
Qy 268 YRLVRYEDLARDPLTVIRELYAFTGLTPO-----LQTWIHNITHGSGPGARRAFKTT 322
Db 412 ----VKEEDLIQPP-----RGARTVDGLRLNRVGVQVYLAALW-----AGSGSVPLYNLMEDA 460
Qy 323 SRDALSVSQAW---RHTLPF---AKIRRVQELCGGALQLLGVRSVHSELEORDLSDLL 375
Db 461 ATAETSRVQNWQLRHGAALDAGGVEVRATPELLA-----RVVEEMARVEAEV--- 509
Qy 376 LPRGMDSPFK 384
Db 510 ---GPDPRF 515

RESULT 15
T50744
spheroidene monooxygenase [imported] - Rhodobacter sphaeroides
N:Alternate names: crtA protein
C:Species: Rhodobacter sphaeroides
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C:Accession: T50744
R:Choudhary, M.; Kaplan, S.
Nucleic Acids Res. 28, 862-867, 2000
A:Title: DNA sequence analysis of the photosynthesis region of Rhodobacter sphaeroides 2.
A:Reference number: Z25222; MUID:20115911; PMID:10648776
A:Accession: T50744
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-327 <CHO>
A:Cross-references: UNIPROT:Q9RFD1; EMBL:AF195122; PIDN:AAF24288.1
A:Experimental source: strain 2.4.1
C:Genetics:
A:Gene: crtA
C:Function:
A:Pathway: carotenoid biosynthesis
C:Superfamily: spheroidene monooxygenase

Query Match 4.5%; Score 92.5; DB 2; Length 327;
Best Local Similarity 21.7%; Pred. No. 3;
Matches 73; Conservative 29; Mismatches 147; Indels 87; Gaps 14;
Qy 9 TVMLSLMLVQTGILVFLVSRQVPSAPGLGERVHVLVLS--WRSGSSFVGLFSQHPDVF 67

```

Db      |||||
3  TVTLSIFRNEFEKRLWVLGQMTANKLGM-----HYLPKAKFWKMFSGSGTQGFTPKEN-- 56
Qy      |||||
68  YLMEPAWHVVDTLGSGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNISDLFQMAVSR 127
Db      |||||
57  -----WHVWSILAW--PDEETARREVAES-----PIYORWTMADES-----T 94
Qy      |||||
128  ALCSPVPCEAFARGNISSEEVCKELCATRPFGLAQEAC-----SSYSHVV-- 172
Db      |||||
95  VLLQPTSAWGKWDGKEPFPV-KPASDVRPIAALTRATVVKFAERFWGARAASHMIGR 153
Qy      |||||
173  LKEVRF-----FNQVLYPLSDPALNRIYVHLVRDPRVLRSEQTAKALARDNGI 224
Db      |||||
154  NKDVVFKIGVGEVPFVQVTFESIWPDAARWR-----SSPRGAGGPHGEAIKAVRAENWF 208
Qy      |||||
225  -----VLGTNGTWEADPRLRVNVVCRSH-----VRIAEALHK 259
Db      |||||
209  KEELYARFQILGTIGKWEKDP-----VGEALTARPSEAPKPAPAPAVAQAPAPAAEPKA 264
Qy      |||||
260  PPPFLODRYRLVRYEDLARDPLTVIRELYAFTGIGL 295
Db      |||||
265  PAPVAEKPALAVEMPKPAEPKPVVEAPKPKPERSRGL 300
```

Search completed: June 23, 2005, 08:50:11
Job time : 11.9593 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.935 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-7
Perfect score: 2057
Sequence: 1 MRLPRFSTWMLSLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues
Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09D_PUBCOMB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2057	100.0	395	16	US-10-697-828-7
2	2057	100.0	395	16	US-10-841-707-8
3	2042.5	99.3	418	9	US-09-927-602-5
4	1727	84.0	395	9	US-09-927-602-3
5	1708	83.0	395	9	US-09-927-602-2
6	1708	83.0	395	15	US-10-258-080-5
7	1708	83.0	395	16	US-10-648-593-159
8	1708	83.0	395	16	US-10-697-828-13
9	1536.5	74.7	390	9	US-09-927-602-4
10	1536.5	74.7	390	16	US-10-697-828-8
11	1333	64.8	394	16	US-10-408-765A-2211

12	1017.5	49.5	386	9	US-09-816-825-2
13	1017.5	49.5	386	13	US-10-007-262-1
14	1015.5	49.4	380	16	US-10-841-707-6
15	986.5	48.0	386	14	US-10-427-631-11
16	712	34.6	169	9	US-09-927-602-6
17	677	32.9	169	9	US-09-927-602-7
18	647.5	31.5	484	14	US-10-212-933-4
19	647.5	31.5	531	9	US-09-833-790-255
20	647.5	31.5	531	16	US-10-755-889-284
21	640	31.1	483	14	US-10-212-933-2
22	636.5	30.9	530	16	US-10-723-860-1409
23	563.5	27.4	479	13	US-10-087-192-126
24	550	26.7	481	13	US-10-087-192-123
25	549	26.7	411	14	US-10-021-660-128
26	549	26.7	411	15	US-10-211-462-97
27	549	26.7	411	16	US-10-408-765A-395
28	549	26.7	411	16	US-10-723-860-1544
29	529.5	25.7	171	9	US-09-927-602-8
30	377	18.3	169	9	US-09-927-602-9
31	351.5	17.1	174	9	US-09-927-602-11
32	338	16.4	179	9	US-09-927-602-10
33	126	6.1	387	14	US-10-126-279-21
34	126	6.1	387	14	US-10-286-606-21
35	126	6.1	387	16	US-10-891-383-21
36	118	5.7	596	16	US-10-697-828-9
37	118	5.7	1212	17	US-10-479-472A-2
38	118	5.7	1222	16	US-10-697-828-15
39	118	5.7	1222	16	US-10-475-446-4
40	117	5.7	1207	16	US-10-697-828-17
41	117	5.7	1207	17	US-10-479-472A-4
42	100.5	4.9	754	15	US-10-425-563-18
43	100.5	4.9	754	18	US-10-972-789A-18
44	97.5	4.7	1261	16	US-10-437-963-200710
45	95.5	4.6	502	15	US-10-369-493-19513

ALIGNMENTS

RESULT 1
US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PRT
; ORGANISM: mus musculus
US-10-697-828-7

Query Match	100.0%;	Score 2057;	DB 16;	Length 395;
Best Local Similarity	100.0%;	Pred. No. 9.4e-196;		
Matches 395;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MRLPRFSTWMLSLMVQTGILVFLVSRQVPSPAGIGERVHVLVLSWSRSGSFVQGLF	60	
Db	1	MRLPRFSTWMLSLMVQTGILVFLVSRQVPSPAGIGERVHVLVLSWSRSGSFVQGLF	60	
Qy	61	SQHPDVFYLMPEAWHVWDTTISQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNRISDL	120	

```
Db 61 SQHPDVFYLMPEAWHVMDTSLQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 395
Db 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 395

RESULT 2
US-10-841-707-8
; Sequence 8, Application US/10841707
; Patent No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; PRIOR FILING DATE: 2004-05-06
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-841-707-8

Query Match 100.0%; Score 2057; DB 16; Length 395;
Best Local Similarity 100.0%; Pred. No. 9.4e-196;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Db 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Qy 61 SQHPDVFYLMPEAWHVMDTSLQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 SQHPDVFYLMPEAWHVMDTSLQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Qy 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
```

```
Db 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 395
Db 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 395

RESULT 3
US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Query Match 99.3%; Score 2042.5; DB 9; Length 418;
Best Local Similarity 99.7%; Pred. No. 2.8e-194;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 60
Db 25 MRLPRFSSTVMSLLMVQGTGILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFVGQLF 84
Qy 61 SQHPDVFYLMPEAWHVMDTSLQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 85 SQHPDVFYLMPEAWHVMDTSLQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 144
Qy 121 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 180
Db 145 FQAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFFN 204
Qy 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 205 LQVLYPLLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 264
Qy 241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 300
Db 265 VVNEVCRSHVRIAE-ALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLGTPQLQ 323
Qy 301 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 360
Db 324 TWIHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYS 383
Qy 361 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 395
Db 384 VHSLEQORDLSLDLLPRGMDSFKWASSTEKOPES 418

RESULT 4
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
```

;; CURRENT FILING DATE: 2001-08-09
;; PRIOR APPLICATION NUMBER: US 09/638,211
;; PRIOR FILING DATE: 2000-08-11
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 395
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: synthetic construct
;; NAME/KEY: VARIANT
;; LOCATION: (1)...(395)
;; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 84.0%; Score 1727; DB 9; Length 395;
Best Local Similarity 84.7%; Pred. No. 6.8e-163;
Matches 333; Conservative 16; Mismatches 44; Indels 0; Gaps 0;
Qy 1 MRLPRFSTVMSLLMVQTGLVFLVSRQVPSPAGLGERVHVLVLSWSRSGSFVQOLF 60
Db 1 MRLPRFSTVXTXLLLAQTLLFLVSRPGSPAGGEXRVHVLVLSWSRSGSFVQOLF 60
Qy 61 SQHPDVFLMEPAWHVMDTSLSQSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNLSDL 120
Db 61 SQHPDVFLMEPAWHVMDTSLSQSAAATLHMAVRDLIRSVFLCDMDVDFDAYLPWRNLSDL 120
Qy 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Qy 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Qy 181 LOVLYPLSDPALNLRVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEAAHKPPFPQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPFPQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
Db 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
Qy 361 VHSLEQRDLSDLLPRGMDSPFKWASSTKQPS 393
Db 361 VYSEDEQRDLSDLLPRGMDSPFKWASSTKQPS 393

RESULT 5
US-09-927-602-2
;; Sequence 2, Application US/09927602
;; Patent No. US20020061562A1
;; GENERAL INFORMATION:
;; APPLICANT: Akama, Michiko N.
;; APPLICANT: Fukuda, Michiko N.
;; TITLE OF INVENTION: Methods of Treating Macular Corneal
;; FILE OF INVENTION: Dystrophy
;; FILE REFERENCE: P-LJ 4852
;; CURRENT APPLICATION NUMBER: US/09/927,602
;; PRIOR FILING DATE: 2001-08-09
;; PRIOR APPLICATION NUMBER: US 09/638,211
;; PRIOR FILING DATE: 2000-08-11
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 2
;; LENGTH: 395
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-09-927-602-2
Query Match 83.0%; Score 1708; DB 9; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;

Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
Qy 1 MRLPRFSTVMSLLMVQTGLVFLVSRQVPSPAGLGERVHVLVLSWSRSGSFVQOLF 60
Db 1 MRLPRVSTAVTALLAQTFLLFLVSRPGSPAGGEXRVHVLVLSWSRSGSFVQOLF 60
Qy 61 SQHPDVFLMEPAWHVMDTSLSQSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRNLSDL 120
Db 61 NQHPDVFLMEPAWHVMDTSLSQSAAATLHMAVRDLIRSVFLCDMDVDFDAYLPWRNLSDL 120
Qy 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARAGNISSEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Qy 181 LOVLYPLSDPALNLRVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRVHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPRLR 240
Qy 241 VVNEVCRSHVRIAEAAHKPPFPQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPFPQDRYRLVRYEDLARDPLTVIRELYAFTGLTLPQLQ 300
Qy 301 TWIHNTHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
Db 301 AWIHNTHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGALQLLGYRS 360
Qy 361 VHSLEQRDLSDLLPRGMDSPFKWASSTKQPS 393
Db 361 VYSEDEQRDLSDLLPRGMDSPFKWASSTKQPS 393
RESULT 6
US-10-258-080-5
;; Sequence 5, Application US/10258080
;; Publication No. US20040029125A1
;; GENERAL INFORMATION:
;; APPLICANT: Incyte Genomics, Inc.
;; APPLICANT: POLICKY, Jennifer L.
;; APPLICANT: HAFALIA, April J.A.
;; APPLICANT: BURFORD, Neil
;; APPLICANT: RING, Huijun Z.
;; APPLICANT: LAL, Preeti G.
;; APPLICANT: TRIBOULEY, Catherine M.
;; APPLICANT: YAO, Monique G.
;; APPLICANT: YUE, Henry
;; APPLICANT: TANG, Y. Tom
;; APPLICANT: ARVIZU, Chandra S.
;; APPLICANT: DAS, Debopriya
;; APPLICANT: SANJANWALA, Madhusudan M.
;; APPLICANT: GANDHI, Aameena R.
;; APPLICANT: REDDY, Roopa M.
;; APPLICANT: KHAN, Farrah A.
;; APPLICANT: BAUGHN, Mariah R.
;; APPLICANT: RAMKUNAR, Jayalaxmi
;; APPLICANT: GRIFFIN, Jennifer K.
;; APPLICANT: AU-YOUNG, Janice K.
;; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES
;; FILE REFERENCE: PI-0070 USN
;; CURRENT APPLICATION NUMBER: US/10/258,080
;; CURRENT FILING DATE: 2002-10-15
;; PRIOR APPLICATION NUMBER: US 60/203,509
;; PRIOR FILING DATE: 2000-05-11
;; PRIOR APPLICATION NUMBER: US 60/202,234
;; PRIOR FILING DATE: 2000-05-05
;; PRIOR APPLICATION NUMBER: US 60/200,185
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/198,403
;; PRIOR FILING DATE: 2000-04-19
;; PRIOR APPLICATION NUMBER: PCT/US01/11869
;; PRIOR FILING DATE: 2001-04-12
;; PRIOR APPLICATION NUMBER: US 60/197,590
;; PRIOR FILING DATE: 2000-04-13
;; NUMBER OF SEQ ID NOS: 20

```
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match      83.0%; Score 1708; DB 15; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVWLSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVSSWRSGSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVLVSSWRSGSFGQLF 60

Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120

Qy 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180

Qy 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240

Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300

Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYRS 360
Db 301 AWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYRS 360

Qy 361 VHSELEQRDLSDLLPRGMDSKWASSTEKQES 395
Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 7
US-10-648-593-159
; Sequence 159, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,395
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 159
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-159

Query Match      83.0%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVWLSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVSSWRSGSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVLVSSWRSGSFGQLF 60
```

```
Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120

Qy 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180

Qy 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240

Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300

Qy 301 TWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYRS 360
Db 301 AWIHNITHGSGPGARREAFKTTSDALSVSQAWRHHTLPFAKIRRVQELCGGALQLLYRS 360

Qy 361 VHSELEQRDLSDLLPRGMDSKWASSTEKQES 395
Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 8
US-10-697-828-13
; Sequence 13, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 395
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-13

Query Match      83.0%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.3e-161;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MRLPRFSTVWLSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVSSWRSGSFGQLF 60
Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGPSPAGGEARVHVLVSSWRSGSFGQLF 60

Qy 61 SQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120
Db 61 NQHPDVFLMEPAWHVMDTLTSGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISDL 120

Qy 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180
Db 121 FQAVSRALCSPPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFN 180

Qy 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db 181 LOVLYPLSDPALNLRIVHLVRDPRAVLRREQTAKALARDNGIVLGTNGTWVEADPRLR 240

Qy 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
Db 241 VVNEVCRSHVRIAEAAHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQLQ 300
```


Qy 301 TWIHNTHSGPGARRAEAFKTTSDALSVSQAWRHTLPFAKIRRVQBLCGALQLGYR 360
 Db 301 AMIHNTHSGPGARRAEAFKTTSSNALNVSQAWRHLPFAKIRRVQBLCGALQLGYR 360
 Qy 361 VHSLEQRDLSDLLPRGMDSPKASSTKQPS 395
 Db 361 VYSEQRNLALDLVLPRLNGFTWASSTASHPRN 395

RESULT 9
 US-09-927-602-4
 ; Sequence 4, Application US/09927602
 ; Patent No. US20020061562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; APPLICANT: Akama, Tomoya O.
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal
 ; Dystrophy
 ; FILE REFERENCE: P-LJ 4852
 ; CURRENT APPLICATION NUMBER: US/09/927,602
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: US 09/638,211
 ; PRIOR FILING DATE: 2000-08-11
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 390
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-09-927-602-4

Query Match 74.7%; Score 1536.5; DB 9; Length 390;
 Best Local Similarity 76.0%; Pred. No. 6.3e-144;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
 Qy 1 MRLPRFSTVMSLLMVQGTG-ILVFLVSRQVPSPPAGLGRVHVLVSSWRSGSFGQL 59
 Db 1 MWLPFSSTKTVTLVLLAQTTCLLLFIISRPSPAGGEDRVHVLVSSWRSGSFGQL 60
 Qy 60 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLRSVFLCDMDVFDALPWRNLS 119
 Db 61 FSQHPDVFLMEPAWHVWDTLSQGSAAATLHMAVRDLMSIFLCDMDVFDAYMPSRLNS 120
 Qy 120 LFQWAVSRALCSPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKEVRF 179
 Db 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLARACRSYSHVVLKEVRF 180
 Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
 Db 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
 Qy 240 RVNVEVCRSHVRIAEAAALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPQL 299
 Db 241 RLIREVCRSHVRIAEAAATLKPPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300
 Qy 300 QTWIHNTHSGPGARRAEAFKTTSDALSVSQAWRHTLPFAKIRRVQBLCGALQLGYR 359
 Db 301 EAWIHNTHSGGIGKPIEAFHTSSNARNVSQAWRHLPFTKILRVQVCGALQLGYR 360
 Qy 360 SVHSELRDLSDLLPRGMDSPKAS 387
 Db 361 PVYSADQQRDLTLVLPGRGDHFSWAS 388

RESULT 10
 US-10-697-828-8
 ; Sequence 8, Application US/10697828
 ; Publication No. US20040185546A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, Steven
 ; APPLICANT: Lee, Jin Kyu
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6

FILE REFERENCE: UCAL-138DIV
 ; CURRENT APPLICATION NUMBER: US/10/697,828
 ; CURRENT FILING DATE: 2003-10-29
 ; PRIOR APPLICATION NUMBER: 09/593,828
 ; PRIOR FILING DATE: 2000-06-13
 ; PRIOR APPLICATION NUMBER: 60/144,694
 ; PRIOR FILING DATE: 1999-07-20
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 8
 ; LENGTH: 390
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-697-828-8
 Query Match 74.7%; Score 1536.5; DB 16; Length 390;
 Best Local Similarity 76.0%; Pred. No. 6.3e-144;
 Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;
 Qy 1 MRLPRFSTVMSLLMVQGTG-ILVFLVSRQVPSPPAGLGRVHVLVSSWRSGSFGQL 59
 Db 1 MWLPFSSTKTVTLVLLAQTTCLLLFIISRPSPAGGEDRVHVLVSSWRSGSFGQL 60
 Qy 60 FSQHPDVFLMEPAWHVWDTLSQGSAPALHMAVRDLRSVFLCDMDVFDALPWRNLS 119
 Db 61 FSQHPDVFLMEPAWHVWDTLSQGSAAATLHMAVRDLMSIFLCDMDVFDAYMPSRLNS 120
 Qy 120 LFQWAVSRALCSPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKEVRF 179
 Db 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLARACRSYSHVVLKEVRF 180
 Qy 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPRL 239
 Db 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
 Qy 240 RVNVEVCRSHVRIAEAAALHKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLTPQL 299
 Db 241 RLIREVCRSHVRIAEAAATLKPPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPQL 300
 Qy 300 QTWIHNTHSGPGARRAEAFKTTSDALSVSQAWRHTLPFAKIRRVQBLCGALQLGYR 359
 Db 301 EAWIHNTHSGGIGKPIEAFHTSSNARNVSQAWRHLPFTKILRVQVCGALQLGYR 360
 Qy 360 SVHSELRDLSDLLPRGMDSPKAS 387
 Db 361 PVYSADQQRDLTLVLPGRGDHFSWAS 388

RESULT 11
 US-10-408-765A-2211
 ; Sequence 2211, Application US/10408765A
 ; Publication No. US20040101874A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Fahy, Eoin D.
 ; APPLICANT: Zhang, Bing
 ; APPLICANT: Gibson, Bradford W.
 ; APPLICANT: Taylor, Steven W.
 ; APPLICANT: Glenn, Gary M.
 ; APPLICANT: Warnock, Dale E.
 ; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
 ; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
 ; FILE REFERENCE: 66088.465
 ; CURRENT APPLICATION NUMBER: US/10/408,765A
 ; CURRENT FILING DATE: 2003-04-04
 ; NUMBER OF SEQ ID NOS: 3077
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2211
 ; LENGTH: 394
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-408-765A-2211

```
Query Match      64.8%; Score 1333; DB 16; Length 394;
Best Local Similarity 78.1%; Pred. No. 1.2e-123;
Matches 257; Conservative 19; Mismatches 47; Indels 6; Gaps 1;

Qy      1 MRLPRFSTWMLSLMVTGTLVFLVSRQVPSSPAGLGERVHVVLSSWRSGSFGQLF 60
Db      1 MMLPRVSTAVTALLAQTEFLFLVSRPGPSSPAGGEARVHVVLSSWRSGSFGQLF 60

Qy      61 SQHPDVFYLMPEPAWHVMDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120
Db      61 NQHPDVFYLMPEPAWHVMDTLSQGSAAALHMAVRDLIRSVFLCDMDVFDAYLPWRRNISDL 120

Qy      121 FQAVSRALCSPVPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKEVRFN 180
Db      121 FQAVSRALCSPVACSAFPRGAISKQDVCKTLCAQPFSLAREACRSYSHVVLKEVRFN 180

Qy      181 LOVLYPLSDPALNLRVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWVEADPRLR 240
Db      181 LOVLYPLSDPALNLRVHLVRDPAVLRVREAGPTILARDNGIVLGTNGKMWVEADPHLR 240

Qy      241 VVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLTLPQL 300
Db      241 LIREVCRSHVRIAEALTKPPPLFGRYLRVRFEDLAREPLAEIRALYFTGLTLPQLE 300

Qy      301 TWIHNTHGSG-----PGARREAFKTT 323
Db      301 AWIHNTHGRSGASQSPSILRLGMRATS 329

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match      49.5%; Score 1017.5; DB 9; Length 386;
Best Local Similarity 56.4%; Pred. No. 3e-92;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy      1 MRLPRFSTWMLSLMVTGTLVLF---LVSRQVPS--SPAGLGERVHVVLSSWRSGSFFV 56
Db      1 MLLPK--KMKLLFLVSQMAILALFFHYSHNSSLKMAQPERMHVVLSSWRSGSFFV 58

Qy      57 GOLFSQHPDVFYLMPEPAWHVMDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
Db      59 GOLFGQHPDVFYLMPEPAWHVMDTFSQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 118

Qy      116 NISDLFQWAVSRALCSPVPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKE 175
Db      119 RQSSLFQWENSALCSPACDIIPQDEIIPRAHCRLLCSQPFVEVVEKACRSYSHVVLKE 178

Qy      176 VRFFNLQVLYPLSDPALNLRVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 234
Db      179 VRFFNLQSLYPLKDPNLNLRVHLVRDPAVLRFRSRTKGDLMISRVIMGHEQKLKK 238

Qy      235 ADPLRLVYVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
Db      239 EQQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLG 297

Qy      295 LTPQLQGTWHTHNGSGPGARREAFKTTSRDALSVQAWRHTLPPFAKIRRVQELCGALQ 354
Db      298 FLPLQGTWHTHTRGKGMG--DHAFTNARDALNSQAWRSLPYEKVSRLOKACGDAMN 355

Qy      355 LLGYRSVHSELEQORDLSLDLL 375
Db      356 LLGYRHRVSRSEQORNLLDLL 376

Query Match      49.5%; Score 1017.5; DB 13; Length 386;
Best Local Similarity 56.4%; Pred. No. 3e-92;
Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy      1 MRLPRFSTWMLSLMVTGTLVLF---LVSRQVPS--SPAGLGERVHVVLSSWRSGSFFV 56
Db      1 MLLPK--KMKLLFLVSQMAILALFFHYSHNSSLKMAQPERMHVVLSSWRSGSFFV 58

Qy      57 GOLFSQHPDVFYLMPEPAWHVMDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
Db      59 GOLFGQHPDVFYLMPEPAWHVMDTFSQSTAWMLHMAVRDLIRAVFLCDMSVFDAYMEPGPR 118

Qy      116 NISDLFQWAVSRALCSPVPVCEAFARGNISSEBCKPLCATRPFGLAQACSSYSHVVLKE 175
Db      119 RQSSLFQWENSALCSPACDIIPQDEIIPRAHCRLLCSQPFVEVVEKACRSYSHVVLKE 178

Qy      176 VRFFNLQVLYPLSDPALNLRVHLVRDPAVLRVREQTAKALARDNGIVLGTNGTWV-E 234
Db      179 VRFFNLQSLYPLKDPNLNLRVHLVRDPAVLRFRSRTKGDLMISRVIMGHEQKLKK 238

Qy      235 ADPLRLVYVNEVCRSHVRIAEALHKPPPLQDRYLRVRYEDLARDPLTVIRELYAFTGLG 294
Db      239 EQQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLG 297

Qy      295 LTPQLQGTWHTHNGSGPGARREAFKTTSRDALSVQAWRHTLPPFAKIRRVQELCGALQ 354
Db      298 FLPLQGTWHTHTRGKGMG--DHAFTNARDALNSQAWRSLPYEKVSRLOKACGDAMN 355

Qy      355 LLGYRSVHSELEQORDLSLDLL 375
Db      356 LLGYRHRVSRSEQORNLLDLL 376

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.3029 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPRFSKTVTLVLLAQT.....LTDLVLPGRGHFMSWSPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US10F_PUBCOMB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*

20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*

21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2051	100.0	390	9	US-09-927-602-4
2	2051	100.0	390	16	US-10-697-828-8
3	1758.5	85.7	395	9	US-09-927-602-3
4	1729.5	84.3	395	9	US-09-927-602-2
5	1729.5	84.3	395	15	US-10-258-080-5
6	1729.5	84.3	395	15	US-10-648-593-159
7	1729.5	84.3	395	16	US-10-697-828-13
8	1536.5	74.9	395	16	US-10-697-828-7
9	1536.5	74.9	395	16	US-10-841-707-8
10	1523	74.3	418	9	US-09-927-602-5
11	1513.5	73.8	394	16	US-10-408-765A-2211

12	1008	49.1	386	9	US-09-816-825-2	Sequence 2, Appli
13	1008	49.1	386	13	US-10-007-262-1	Sequence 1, Appli
14	1003.5	48.9	380	16	US-10-841-707-6	Sequence 6, Appli
15	970	47.3	386	14	US-10-427-631-11	Sequence 11, Appli
16	808	39.4	169	9	US-09-927-602-7	Sequence 7, Appli
17	755	36.8	169	9	US-09-927-602-6	Sequence 6, Appli
18	665	32.4	483	14	US-10-212-933-2	Sequence 2, Appli
19	659.5	32.2	530	16	US-10-723-860-1409	Sequence 1409, Ap
20	656.5	32.0	484	14	US-10-212-933-4	Sequence 4, Appli
21	656.5	32.0	531	9	US-09-833-790-255	Sequence 255, App
22	656.5	32.0	531	16	US-10-755-889-284	Sequence 284, App
23	579.5	28.3	479	13	US-10-087-192-126	Sequence 126, App
24	576	28.1	481	13	US-10-087-192-123	Sequence 123, App
25	514.5	25.1	171	9	US-09-927-602-8	Sequence 8, Appli
26	513.5	25.0	411	14	US-10-021-660-128	Sequence 128, App
27	513.5	25.0	411	15	US-10-211-462-97	Sequence 97, Appl
28	513.5	25.0	411	16	US-10-408-765A-395	Sequence 395, App
29	513.5	25.0	411	16	US-10-723-860-1544	Sequence 1544, Ap
30	372	18.1	169	9	US-09-927-602-9	Sequence 9, Appli
31	351.5	17.1	174	9	US-09-927-602-11	Sequence 11, Appl
32	335	16.3	179	9	US-09-927-602-10	Sequence 10, Appl
33	119.5	5.8	596	16	US-10-697-828-9	Sequence 9, Appli
34	119.5	5.8	1212	17	US-10-479-472A-2	Sequence 2, Appli
35	119.5	5.8	1222	16	US-10-697-828-15	Sequence 15, Appl
36	119.5	5.8	1222	16	US-10-475-446-4	Sequence 4, Appli
37	112.5	5.5	1207	16	US-10-697-828-17	Sequence 17, Appl
38	112.5	5.5	1207	17	US-10-479-472A-4	Sequence 4, Appli
39	105	5.1	359	14	US-10-411-976-11	Sequence 11, Appl
40	102.5	5.0	387	14	US-10-126-279-21	Sequence 21, Appl
41	102.5	5.0	387	14	US-10-286-606-21	Sequence 21, Appl
42	102.5	5.0	387	16	US-10-891-383-21	Sequence 21, Appl
43	97	4.7	299	14	US-10-156-761-9846	Sequence 9846, Ap
44	95	4.6	885	17	US-10-732-923-6830	Sequence 6830, Ap
45	94.5	4.6	2588	16	US-10-437-963-138279	Sequence 138279,

ALIGNMENTS

RESULT 1
US-09-927-602-4
; Sequence 4, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 100.0%; Score 2051; DB 9; Length 390;
Beat Local Similarity 100.0%; Pred. No. 2.4e-204;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MWLPFRFSKTVTLVLLAQTCLLIIFISRRGPSPAGGEDRVHVLVSSWFGSFLGQL	60
Db	1	MWLPFRFSKTVTLVLLAQTCLLIIFISRRGPSPAGGEDRVHVLVSSWFGSFLGQL	60
Qy	61	FSQHPDVFYLMERPAHWVTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSRNL	120
Db	61	FSQHPDVFYLMERPAHWVTTLSQGSAAATLHMAVRDLMRSLFLCDMDVFDAYMPSRNL	120

```
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 180
QY 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 240
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
QY 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
DB 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390

RESULT 2
US-10-697-828-8
; Sequence 8, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 390
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-8

Query Match 100.0%; Score 2051; DB 16; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.4e-204;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
DB 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
QY 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYMPOSRNLSA 120
DB 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYMPOSRNLSA 120
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 180
QY 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 240
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
```

```
QY 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390
DB 361 PVSADQQRDLTDLVLPRGPDHFSWASPD 390

RESULT 3
US-09-927-602-3
; Sequence 3, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
; NAME/KEY: VARIANT
; LOCATION: (1)...(395)
; OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 85.7%; Score 1758.5; DB 9; Length 395;
Best Local Similarity 87.1%; Pred. No. 6.9e-174;
Matches 338; Conservative 10; Mismatches 39; Indels 1; Gaps 1;

QY 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 60
DB 1 MWLPRFSSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVLSWRSGSFLGQL 59
QY 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYMPOSRNLSA 120
DB 61 FSQHPDVFLMEPAWHVMTTLSQGSAAATLHMAVRDLMRSLFCLDMDFDAYMPOSRNLSA 119
QY 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 180
DB 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVREF 179
QY 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 240
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSRAAGFILARDNGIVLGTNGKWEADPHL 239
QY 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 300
DB 241 RLIREVCRSHVRIAEATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFTGLTLPOL 299
QY 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 360
DB 301 EAWIHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPFTKILRQVEVCAGALQLLGYR 359
QY 361 PVSADQQRDLTDLVLPRGPDHFSWAS 388
DB 361 PVSADQQRDLTDLVLPRGPDHFSWAS 387

RESULT 4
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
```

```

; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927.602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
; ORIGIN: 927-602-2
US-09-927-602-2

Query Match      84.3%; Score 1729.5; DB 9; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRFSKTVTVLLLAQTCLLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db 1 MWLPRVSSSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLVLSWRSGSFLGQL 59

Qy 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDYMPOSRLSA 120
Db 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCMDVDFDYLPMWRRLSD 119

Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSAECVKPLCARQSFTLAREACRSYSHVVLKEVRF 179

Qy 181 NLQVLYPLSDPALNLRIHVLVRDPAVLRSREAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIHVLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239

Qy 241 RLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPOL 300
Db 240 RVREVCRSVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPOL 299

Qy 301 EAWHNTTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWHNTTHGSGGPGARAEAFKTSRNALNVSOQWRHALPFAKIRRVQELCAGALQLLGYR 359

Qy 361 PVYSADQORDLTLDLVLPRGPDHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

RESULT 5
US-10-258-080-5
; Sequence 5, Application US/10258080
; Publication No. US20040029125A1
; GENERAL INFORMATION:
; APPLICANT: Incyte Genomics, Inc.
; APPLICANT: POLICKY, Jennifer L.
; APPLICANT: HAPALIA, April J.A.
; APPLICANT: BURFORD, Neil
; APPLICANT: RING, Huijun Z.
; APPLICANT: LAL, Freeti G.
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: YAO, Monique G.
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: ARVIZU, Chandara S.
; APPLICANT: DAS, Debopriya
; APPLICANT: SANJANWALA, Madhusudan M.
; APPLICANT: GANDHI, Aneena R.
; APPLICANT: REDDY, Roopa M.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: AU-YOUNG, Janice K.
; TITLE OF INVENTION: DRUG METABOLIZING ENZYMES

```

```

; FILE REFERENCE: PI-0070 USN
; CURRENT APPLICATION NUMBER: US/10/258,080
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/203,509
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/202,234
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/200,185
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/198,403
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: PCT/US01/11869
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/197,590
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029125A1 7472777CD1
US-10-258-080-5

Query Match      84.3%; Score 1729.5; DB 15; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRFSKTVTVLLLAQTCLLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
Db 1 MWLPRVSSSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLVLSWRSGSFLGQL 59

Qy 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDYMPOSRLSA 120
Db 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCMDVDFDYLPMWRRLSD 119

Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPACSAFPRGAISSAECVKPLCARQSFTLAREACRSYSHVVLKEVRF 179

Qy 181 NLQVLYPLSDPALNLRIHVLVRDPAVLRSREAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIHVLVRDPAVLRSREQTAKALARDNGIVLGTNGTWVEADPGL 239

Qy 241 RLIREVCRSHVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPOL 300
Db 240 RVREVCRSVRIAEATLKPPLGRYRLVRFEDLAREPLAEIRALYAFGLTLPOL 299

Qy 301 EAWHNTTHGSGIGKPIEAFTSSRNARNVSOQWRHALPFTKILRVOECAGALQLLGYR 360
Db 300 EAWHNTTHGSGGPGARAEAFKTSRNALNVSOQWRHALPFAKIRRVQELCAGALQLLGYR 359

Qy 361 PVYSADQORDLTLDLVLPRGPDHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

RESULT 6
US-10-648-593-159
; Sequence 159, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; PRIOR FILING DATE: 2002-08-27

```

;
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 159
; LENGTH: 395
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-648-593-159

Query Match 84.3%; Score 1729.5; DB 16; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
Db 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 59
Qy 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYMPQSRNLSA 120
Db 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYLPWRNLS 119
Qy 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGTWEADPGL 239
Qy 241 RLIREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 300
Db 240 RVREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 299
Qy 301 EAWIHNTGSGICKPIEAFTSSRNARNVSQAWRHALPFTKILRVOEVCAGALQLLGYR 360
Db 300 EAWIHNTGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYR 359
Qy 361 PVYSADQORDLTDLVLPGRPDHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRLNGFTWAS 387

RESULT 7

US-10-697-828-13
; Sequence 13, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 395
; TYPE: PR1
; ORGANISM: homo sapiens
US-10-697-828-13

Query Match 84.3%; Score 1729.5; DB 16; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.2e-171;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
Db 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 59

Qy 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYMPQSRNLSA 120
Db 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYLPWRNLS 119
Qy 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGTWEADPGL 239
Qy 241 RLIREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 300
Db 240 RVREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 299
Qy 301 EAWIHNTGSGICKPIEAFTSSRNARNVSQAWRHALPFTKILRVOEVCAGALQLLGYR 360
Db 300 EAWIHNTGSGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYR 359
Qy 361 PVYSADQORDLTDLVLPGRPDHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRLNGFTWAS 387

RESULT 8

US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PR1
; ORGANISM: mus musculus
US-10-697-828-7

Query Match 74.9%; Score 1536.5; DB 16; Length 395;
Best Local Similarity 76.0%; Pred. No. 8.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 60
Db 1 MWLPRESSKTVTLIAQTCLLFIISRPSPAGGEDRVHVLVLSWSSGSSFLGQL 59
Qy 61 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYMPQSRNLSA 120
Db 60 FSOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVFDAYLPWRNLS 119
Qy 121 FFWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEAAAGPILARDNGIVLGTNGTWEADPGL 239
Qy 241 RLIREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 300
Db 240 RVREVCRSHVRIAEATLKPPFLGRYRLVRFEDELAREPLAEIRALYAFTGLTTPQL 299

Qy 301 EAWHNTHTSGGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
Db 300 QTWIHNTHTSGGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 359
Qy 361 PVYSADOQRDLTLDLVLPRGDPHFSWAS 388
Db 360 SVHSELEQRDLSDLLLPGRGMDSPFKWAS 387

RESULT 9

US-10-841-707-8
; Sequence 8, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-06
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-841-707-8

Query Match 74.9%; Score 1536.5; DB 16; Length 395;
Best Local Similarity 76.0%; Pred. No. 8.9e-151;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

Qy 1 MWLPFSSKTVTLAAQTCLLLFIISRPSPGAGGEDRVHVLVLSWRSFGSLGQL 60
Db 1 MRLPFSSTVMLSLLMVQTG-ILVFLVSRQVPSPAGLGERVHVLVLSWRSFGSLGQL 59
Qy 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVARDLMSIFLCMDVDFDAYMPSQSNLSA 120
Db 60 FSOHPDVFYLMPEAWHVTTLTSGSAAPALHMAVRDLIRSFLCDMDVDFDAYLPWRRNISD 119
Qy 121 FFWNATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
Db 120 LFQWAVSRALCSPVCEAFARGNISSEEVCKPLCATRPFGLAQACRSYSHVVLKEVRF 179
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWEADPHL 239
Qy 241 RLIREVCRSHVRIAEEATLKPFPFLGRYLRVRFEDLAREPLAEIRALYFTGLTLPQL 300
Db 240 RVNEVCRSHVRIAEEALHKPPFLQDRYLRVRFEDLARDPLTVIRELYAFTGLTLPQL 299
Qy 301 EAWHNTHTSGGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
Db 300 QTWIHNTHTSGGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 359
Qy 361 PVYSADOQRDLTLDLVLPRGDPHFSWAS 388
Db 360 SVHSELEQRDLSDLLLPGRGMDSPFKWAS 387

RESULT 10

US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US2002006152A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.

; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Query Match 74.3%; Score 1523; DB 9; Length 418;
Best Local Similarity 75.8%; Pred. No. 2.5e-149;
Matches 294; Conservative 30; Mismatches 62; Indels 2; Gaps 2;
Qy 1 MWLPFSSKTVTLAAQTCLLLFIISRPSPGAGGEDRVHVLVLSWRSFGSLGQL 60
Db 25 MRLPFSSTVMLSLLMVQTG-ILVFLVSRQVPSPAGLGERVHVLVLSWRSFGSLGQL 83
Qy 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMVARDLMSIFLCMDVDFDAYMPSQSNLSA 120
Db 84 FSOHPDVFYLMPEAWHVTTLTSGSAPALHMAVRDLIRSFLCDMDVDFDAYLPWRRNISD 143
Qy 121 FFWNATSRALCSPACSPAPRGITISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRF 180
Db 144 LFQWAVSRALCSPVCEAFARGNISSEEVCKPLCATRPFGLAQACRSYSHVVLKEVRF 203
Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL 240
Db 204 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSREQTAKALARDNGIVLGTNGTWEADPHL 263
Qy 241 RLIREVCRSHVRIAEEATLKPFPFLGRYLRVRFEDLAREPLAEIRALYFTGLTLPQL 300
Db 264 RVNEVCRSHVRIAEEALH-KPPFPFLQDRYLRVRFEDLARDPLTVIRELYAFTGLTLPQL 322
Qy 301 EAWHNTHTSGGKPIEAFTSSRNARNVSQAWRHLPFTKILRVOECAGALQLLGYR 360
Db 323 QTWIHNTHTSGGPGARREAFKTTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQLLGYR 382
Qy 361 PVYSADOQRDLTLDLVLPRGDPHFSWAS 388
Db 383 SVHSELEQRDLSDLLLPGRGMDSPFKWAS 410

RESULT 11

US-10-408-765A-2211
; Sequence 2211, Application US/10408765A
; Publication No. US2004010187A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2211
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2211

```
Query Match 73.8%; Score 1513.5; DB 16; Length 394;
Best Local Similarity 92.6%; Pred. No. 2.2e-148;
Matches 289; Conservative 7; Mismatches 15; Indels 1; Gaps 1;

Qy 1 MWLPRFSKVTYVULLAQTCLLLFIISRCGPSPAGGEDRVHVLVLSWRSGSSFLGQL 60
Db 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRFGPSPAGGEARVHVLVLSWRSGSSFLGQL 59
Qy 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLDCMDVDFDAYMPSQSNLSA 120
Db 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLDCMDVDFDAYLPWRRLSD 119
Qy 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPACSAFPRGAIKQDVCKTCTCARQPFSLAREACRSYSHVVLKEVRFF 179
Qy 181 NLQVLYPELLSDPALNLRIHVLRDPRAVLRSREAGPIIARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPELLSDPALNLRIHVLRDPRAVLRSREAGPIIARDNGIVLGTNGKWEADPHL 239
Qy 241 RLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTLPOL 300
Db 240 RLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTLPOL 299
Qy 301 EAWIHNITHGSG 312
Db 300 EAWIHNITHGSG 311

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 49.1%; Score 1008; DB 9; Length 386;
Best Local Similarity 52.3%; Pred. No. 9.1e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSSFLGQ 59
Db 1 MLLPKOKKLLFLVSQMAILALFFHMYSHNISLSMKAQPERHVLVLSWRSGSSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLDCMDVDFDAYM-POSRL 118
Db 61 LFGQHPDVFYLMPEPAWHVWTTKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGPRQ 120
Qy 119 SAFNWTASRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSPACDIIPQDEIIPRAHCRLLCQQPFVEVVEKACRSYSHVVLKEVR 180
Qy 179 FENLOVLYPLLSDPALNLRIHVLRDPRAVLRSREAGPIIARDNGIVLGTN-GKWVEAD 237
Db 181 FENLOVLYPLLKDPSSLNLHIVLRDPRAVFRSRTKGDLMIDRSIVMGQHQKLUKED 240
Qy 238 PHRLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICSQOLEIYK--TIQSLPKAQERVLLVRYEDLAPAPVAQTSRMVEFVGLGF 298
Qy 297 TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTTKILRVQEVCAQALQL 356
Db 299 LPHLQVWVNIITRGKMGD--HAFHTNARDALNVSOAWRWSLPEYKVSRLQKACGDMNL 356
Qy 357 LGYRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRHRVRSQEQRNLLDLL-----STWTVPE 383

RESULT 13
US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRF
; ORGANISM: H. sapiens
US-10-007-262-1

Query Match 49.1%; Score 1008; DB 13; Length 386;
Best Local Similarity 52.3%; Pred. No. 9.1e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSSPAGGEDRVHVLVLSWRSGSSFLGQ 59
Db 1 MLLPKOKKLLFLVSQMAILALFFHMYSHNISLSMKAQPERHVLVLSWRSGSSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLDCMDVDFDAYM-POSRL 118
Db 61 LFGQHPDVFYLMPEPAWHVWTTKQSTAWMLHMAVRDLIRAVFLCDMSVDFDAYMEPGPRQ 120
Qy 119 SAFNWTASRALCSPACSAFPRGTISKQDVCKTCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSPACDIIPQDEIIPRAHCRLLCQQPFVEVVEKACRSYSHVVLKEVR 180
Qy 179 FENLOVLYPLLSDPALNLRIHVLRDPRAVLRSREAGPIIARDNGIVLGTN-GKWVEAD 237
Db 181 FENLOVLYPLLKDPSSLNLHIVLRDPRAVFRSRTKGDLMIDRSIVMGQHQKLUKED 240
Qy 238 PHRLIREVCRSHVRIAEEATLKP-PPFLRGYRLVRPFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICSQOLEIYK--TIQSLPKAQERVLLVRYEDLAPAPVAQTSRMVEFVGLGF 298
Qy 297 TPQLEAWIHNITHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTTKILRVQEVCAQALQL 356
Db 299 LPHLQVWVNIITRGKMGD--HAFHTNARDALNVSOAWRWSLPEYKVSRLQKACGDMNL 356
Qy 357 LGYRPVYSADQORDLTLDLVLPRGPDHFSWASPD 390
Db 357 LGYRHRVRSQEQRNLLDLL-----STWTVPE 383

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
```

APPLICANT: Yeh, Jiunn-Chern
APPLICANT: Hiraoka, Nobuyoshi
TITLE OF INVENTION: Identification of the Meca-79 Antigen
TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
TITLE OF INVENTION: Conditions
FILE REFERENCE: P-LJ 4149
CURRENT APPLICATION NUMBER: US/10/841.707
CURRENT FILING DATE: 2004-05-06
PRIOR APPLICATION NUMBER: US/09/569,320A
PRIOR FILING DATE: 2000-05-11
NUMBER OF SEQ ID NOS: 29
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 380
TYPE: PRT
ORGANISM: Homo Sapien
US-10-841-707-6

Query Match 48.9%; Score 1003.5; DB 16; Length 380;
Best Local Similarity 53.1%; Pred. No. 2.6e-95;
Matches 206; Conservative 55; Mismatches 102; Indels 25; Gaps 8;

Qy 13 VLLAQTCLLLF-----IISRPSSPAGGEDRVHVLVLSWRSGSFLGQLFSQHP 65
Db 5 LFLVSQMAILALFFHMYSHNISLSMKAQP-----ERMHVLSWRSGSFLGQLFSQHP 60
Qy 66 DVFLMEPAHWMTTLSQGSAAATLHMVARDLMSIFLCMDVDFDAYM-PQSRNLSAFNW 124
Db 61 DVFLMEPAHWMTTQKSTAWMLHMVARDLIRAVFLCDMSVFDAYMEFGPRQSSLFQW 120
Qy 125 ATRALCSPACSAAPPRTGISQDVCKTLCTROPFSLAREACRSYSHVVLKEVFFNLQV 184
Db 121 ENSRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVVEKACRSYSHVVLKEVFFNLQ 180
Qy 185 LYPLLSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEADPHRLI 243
Db 181 LYPLLKPSLNLHIVHVRDPAVRSRRTKGDLMIDSRVIMGQHEQKKEKQPYVYM 240
Qy 244 REVCRSHVRIAEATLKP-PPLGRVRLVRFEDLAREPLAEIRALYAFTGLTLPOL 302
Db 241 QVICOSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLFPHLQT 298
Qy 303 WIHNITGSGIGKPIEAFHTSSRNARNVSQAWRHALPTKILRVQEVCAQALQLLYRPV 362
Db 299 VWNITRCKGMD--HAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNLLGYRHV 356

Qy 363 YSADQORLTLDLVLRPGDPHFSPD 390
Db 357 RSEQORNLDDL-----STWTVPE 377

RESULT 15

US-10-427-631-11
Sequence 11, Application US/10427631
Publication No. US20030175923A1
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; TANG, Y. TOM;
APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
APPLICANT: BAUGHN, Maria R.; LAL, Preeti G.;
APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PP-0592-1 DIV
CURRENT APPLICATION NUMBER: US/10/427,631
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 09/786,240
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: PCT/US99/20989
PRIOR FILING DATE: 1999-09-09
PRIOR APPLICATION NUMBER: US 60/172,220
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: US 60/155,248
PRIOR FILING DATE: 1998-11-04

PRIOR APPLICATION NUMBER: US 60/133,642
PRIOR FILING DATE: 1999-05-11
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PERL Program
SEQ ID NO 11
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match 47.3%; Score 970; DB 14; Length 386;
Best Local Similarity 51.3%; Pred. No. 8.3e-92;
Matches 202; Conservative 52; Mismatches 112; Indels 28; Gaps 7;

Qy 14 LLLAQTCLLLFIISRP-----GPSPPAGGEDRVHVLVLSWRSGSFLGQ 59
Db 1 MLLPKMKLLFLVSQMAILALFFHMYSHNISLSMKAQPERMHVLSWRSGSFLVGQ 60
Qy 60 LFSQHPDVFYLMPEPAHWMTTLSQGSAAATLHMVARDLMSIFLCMDVDFDAYM-PQSRNL 118
Db 61 LFGQHPDVFYLMPEPAHWMTTQKSTAWMLHMVARDLIRAVFLCDMSVFDAYMEGPRRQ 120
Qy 119 SAFPNWATSRALCSPACSAAPPRTGISQDVCKTLCTROPFSLAREACRSYSHVVLKEVR 178
Db 121 SSLFQWENSRALCSAPACDIIPQDESSPGLTAGSCAVNSPLKLEKACRSYSHVVLKEVR 180
Qy 179 FPNLOVLYPLSDPALNLRVHVRDPAVLRSREAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 FPNLQSLYPLKPSLNLHIVHVRDPAVRSRRTKGDLMIDSRVIMGQHEQKKEKBD 240
Qy 238 PHRLIREVCRSHVRIAEATLKP-PPLGRVRLVRFEDLAREPLAEIRALYAFTGLTL 296
Db 241 QPYVMQVICOSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGL 298
Qy 297 TPQLEAWIHNTITGSGIGKPIEAFHTSSRNARNVSQAWRHALPTKILRVQEVCAQALQL 356
Db 299 LPHLQTVWNITRCKGMD--HAFHTNARDALNVSQAWRWSLPYEKVSRLQKACGDAMNL 356

Qy 357 LGYRPVVSADQORLTLDLVLRPGDPHFSPD 390
Db 357 LGYRHRVRSQEQORNLLDDL-----STWTVPE 383

Search completed: June 23, 2005, 09:27:53
Job time : 50.3029 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.6998 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-8

Perfect score: 2051

Sequence: 1 MWLPRFSKTVTVLLAQTT.....LTLDLVLPRGDFHFSWASPD 390

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:**

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:**

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:**

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:**

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:**

5: /cgn2_6/ptodata/1/iaa/FCRUS_COMB.pep:**

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2051	100.0	390	4	US-09-949-016-6813 Sequence 6813, Ap
2	2051	100.0	431	4	US-09-949-016-6893 Sequence 8893, Ap
3	1729.5	84.3	395	4	US-09-949-016-7011 Sequence 7011, Ap
4	1008	49.1	386	3	US-09-045-284A-2 Sequence 2, Appli
5	1008	49.1	386	3	US-09-190-911-1 Sequence 1, Appli
6	970	47.3	386	4	US-09-786-240-11 Sequence 11, Appl
7	665	32.4	483	3	US-09-263-023-2 Sequence 2, Appli
8	665	32.4	483	4	US-09-471-867-2 Sequence 2, Appli
9	659.5	32.2	608	4	US-09-949-016-9449 Sequence 9449, Ap
10	656.5	32.0	484	3	US-09-263-023-4 Sequence 4, Appli
11	656.5	32.0	484	4	US-09-471-867-4 Sequence 4, Appli
12	656.5	32.0	531	4	US-09-949-016-6471 Sequence 6471, Ap
13	577.5	28.2	479	4	US-08-899-514-2 Sequence 2, Appli
14	540.5	26.4	458	2	US-08-655-878-2 Sequence 2, Appli
15	513.5	25.0	411	3	US-09-015-188-2 Sequence 2, Appli
16	105	5.1	359	3	US-09-150-133-11 Sequence 11, Appl
17	105	5.1	359	3	US-09-150-141-11 Sequence 11, Appl
18	105	5.1	359	3	US-09-374-493-11 Sequence 11, Appl
19	105	5.1	359	3	US-09-374-824-11 Sequence 11, Appl
20	105	5.1	359	3	US-09-374-492-11 Sequence 11, Appl
21	105	5.1	359	4	US-09-785-343-11 Sequence 11, Appl
22	105	5.1	359	4	US-10-411-976-11 Sequence 11, Appl
23	97	4.7	199	4	US-09-252-901A-28497 Sequence 28497, A
24	91.5	4.5	242	4	US-09-248-796A-20943 Sequence 9, Appli
25	90.5	4.4	380	3	US-09-150-133-9 Sequence 9, Appli
26	90.5	4.4	380	3	US-09-150-141-9 Sequence 9, Appli
27	90.5	4.4	380	3	US-09-374-493-9 Sequence 9, Appli

28	90.5	4.4	380	3	US-09-374-824-9	Sequence 9, Appli
29	90.5	4.4	380	3	US-09-374-492-9	Sequence 9, Appli
30	90.5	4.4	380	4	US-09-785-343-9	Sequence 9, Appli
31	90.5	4.4	380	4	US-10-411-976-9	Sequence 9, Appli
32	90.5	4.4	984	2	US-08-673-789-9	Sequence 9, Appli
33	90	4.4	575	4	US-09-252-901A-32986	Sequence 32986, A
34	88	4.3	288	4	US-09-818-780-17	Sequence 17, Appl
35	88	4.3	268	4	US-09-818-780-94	Sequence 94, Appl
36	87	4.2	255	4	US-09-902-540-15692	Sequence 15692, A
37	86.5	4.2	376	3	US-09-150-133-7	Sequence 7, Appli
38	86.5	4.2	376	3	US-09-150-141-7	Sequence 7, Appli
39	86.5	4.2	376	3	US-09-374-493-7	Sequence 7, Appli
40	86.5	4.2	376	3	US-09-374-824-7	Sequence 7, Appli
41	86.5	4.2	376	3	US-09-374-492-7	Sequence 7, Appli
42	86.5	4.2	376	4	US-09-785-343-7	Sequence 7, Appli
43	86.5	4.2	376	4	US-10-411-976-7	Sequence 7, Appli
44	86.5	4.2	377	3	US-09-150-133-5	Sequence 5, Appli
45	86.5	4.2	377	3	US-09-150-141-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1

US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match 100.0%; Score 2051; DB 4; Length 390;
Best Local Similarity 100.0%; Pred. No. 2.5e-217;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MWLPRFSKTVTVLLAQTTCLLLFIISRPSPSPAGDEDRVHVLVLSWSGSSFLQOL	60
Db	1	MWLPRFSKTVTVLLAQTTCLLLFIISRPSPSPAGDEDRVHVLVLSWSGSSFLQOL	60
Qy	61	FSQHPDVFYLMPEAHVHTTSLQSSAATLHMAVRDLARSIFCLDMDFDAMPOSRNL	120
Db	61	FSQHPDVFYLMPEAHVHTTSLQSSAATLHMAVRDLARSIFCLDMDFDAMPOSRNL	120
Qy	121	FFWATSRALCSPPACSAFPRGTISKQDVCKLTCTROPFSLAREACRSYSHVLKEVRFF	180
Db	121	FFWATSRALCSPPACSAFPRGTISKQDVCKLTCTROPFSLAREACRSYSHVLKEVRFF	180
Qy	181	NLOVLYPLLSDDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL	240
Db	181	NLOVLYPLLSDDPALNLRIVHLVRDPRAVLRSREAGPILARDNGIVLGTNGKWEADPHL	240
Qy	241	RLIREVCRSHVRIAEEAATLKPPLRCGRVLRVREPLAREPLAEIRALYATFTGLTLPOL	300
Db	241	RLIREVCRSHVRIAEEAATLKPPLRCGRVLRVREPLAREPLAEIRALYATFTGLTLPOL	300
Qy	301	EAWITHNTHSGIGKPIEAFHTSSRNARNVSQAWRHALPFTTKILRVQEVCAQALQLGYR	360

```
|||||EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Db 301
Qy 361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390
Db 361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390

RESULT 2
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20-237,768
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8893

Query Match 100.0%; Score 2051; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 2.9e-217;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MMLPRFSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVSSWRSGSFLGQL 60
Db 42 MMLPRFSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVSSWRSGSFLGQL 101

Qy 61 FSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPSQSNLSA 120
Db 102 FSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPSQSNLSA 161

Qy 121 FFWNATSRALCSPSPACSAFPRGTISKDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 162 FFWNATSRALCSPSPACSAFPRGTISKDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 221

Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 222 NLQVLYPLSDPALNLRIVHLVRDPRAVLSREAAAGPILARDNGIVLGTNGKWEADPHL 281

Qy 241 RLIREVCRSHVRIAEATLKPPLFLGRYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 300
Db 282 RLIREVCRSHVRIAEATLKPPLFLGRYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 341

Qy 301 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Db 342 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 401

Qy 361 PVSADQOQDLTLDLVLPRGPDHFSWASPD 390
Db 402 PVSADQOQDLTLDLVLPRGPDHFSWASPD 431

RESULT 3
US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

```
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7011

Query Match 84.3%; Score 1729.5; DB 4; Length 395;
Best Local Similarity 85.8%; Pred. No. 7.4e-182;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MMLPRFSKTVTVLLLAQTTCLLLFIISRPSPSPAGGEDRVHVLVSSWRSGSFLGQL 60
Db 1 MMLPRVSTAVTALLAQ-TFLLFLVSRPSPSPAGGEARVHVLVSSWRSGSFLGQL 59

Qy 61 FSQHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLMSIFLCMDVDFDAYMPSQSNLSA 120
Db 60 FNOHPDVFLMEPAWHVWTTLSQSSAATLHMAVRDLVRSVFLCDMDVDFDAYLPWRRLSD 119

Qy 121 FFWNATSRALCSPSPACSAFPRGTISKDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 180
Db 120 LQWAVSRALCSPSPACSAFPRGAISSSEAVCKPCARQSFILAREACRSYSHVVLKEVRFF 179

Qy 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWEADPGL 239

Qy 241 RLIREVCRSHVRIAEATLKPPLFLGRYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 300
Db 240 RVREVCRRSHVRIAEATLKPPLFLGRYRLVRFPEDLAREPLAEIRALYAFTGLTLPQL 299

Qy 301 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 360
Db 300 EAMHNTHSGGIGKPIEAFTSSRNARNVQAWRHALPTKILRVOECAGALQLLYR 359

Qy 361 PVSADQOQDLTLDLVLPRGPDHFSWAS 388
Db 360 PVYSEDEQRNALDLVLPRLNGFTWAS 387

RESULT 4
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.1%; Score 1008; DB 3; Length 386;
Best Local Similarity 52.3%; Pred. No. 2.7e-102;
```


; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9449
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-949-016-9449

Query Match 32.2%; Score 659.5; DB 4; Length 608;
 Best Local Similarity 39.1%; Pred. No. 1.5e-63;
 Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

Qy 31 GPSPAG-----CEDRVHVLSSWSRSGSFLQOLFSDHFDVFLMEPAWHVMTTL 81
 Db 223 GVAAPGNGTRGTGGVGDQRQLVVTFTWRSRSGSFFGELFNQNPVEVFLYEPVHVMQKL 282
 Qy 82 SQSAATLHNAVRLMRSIFLCMDVFDAYMPQ---SRNLS--AFFNWTSRALCSPAC 136
 Db 283 YPGDAVSLOGAARDMLSALYRCDSLVSFQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 342
 Qy 137 SAPPRGTISKQD--VCKTCLTRQPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPAL 194
 Db 343 PAYRKEVGLVDVCKK-CPQRLARFEBCRYRTLVIKGVVFDVAVLAPLLRDPAL 401
 Qy 195 NLRVHLVRDPRVLR-----REAGPILARD-----NGIVLTNGKWV 234
 Db 402 DLKVIHLVRDPRVLRSSIRSRHGLRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGV 461
 Qy 235 --EADPH-LRLIREVCHSHVRIAEATLKPPFLGRVRLVRFPEDLAREPLAEIRALYAF 291
 Db 462 GGPADYHALGAMEVICNSMAKTLQTA-LQPPDMWQGHYLVVRYEDLVGDPVKTLLRYDYF 520
 Qy 292 TGLTLTPQLEAHINITHGSG-IGKPIEAHTSSRNARNVSOAQRHALPFTKILRVOEVC 350
 Db 521 VGLLVSPMEQFALNMTSGSSSKP---FVVSARNATQANAWRTALTFOQIKQVEEFC 577
 Qy 351 AGALQLLGYRPPVSADQORDTLDLV 376
 Db 578 YQNAVLYGVRVNSPEVKLSKTL 603

RESULT 10
 US-09-263-023-4
 ; Sequence 4, Application US/09263023
 ; Patent No. 6037159
 ; GENERAL INFORMATION:
 ; APPLICANT: Uchimura, Kenji
 ; APPLICANT: Muramatsu, Hideki
 ; APPLICANT: Kadomatsu, Kenji
 ; APPLICANT: Kannagi, Reiji
 ; APPLICANT: Habuchi, Osami
 ; APPLICANT: Muramatsu, Takashi
 ; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
 ; FILE REFERENCE: TOYAMA1.001AUS
 ; CURRENT APPLICATION NUMBER: US/09/263,023
 ; EARLIER FILING DATE: 1998-03-05
 ; EARLIER FILING DATE: 1998-03-05
 ; EARLIER FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-263-023-4

Query Match 32.0%; Score 656.5; DB 3; Length 484;
 Best Local Similarity 39.7%; Pred. No. 2.3e-63;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

Qy 34 SPAGGEDRVH-VLVLSWSRSGSFLQOLFSDHFDVFLMEPAWHVMTTL 92
 Db 110 APEGVGDKRHMVYVFTWRSRSGSFFGELFNQNPVEVFLYEPVHVMQKLYPGDAVSLOGA 169
 Qy 93 VRDLMSIFLCMDVFDAYMPQ---SRNLS--AFFNWTSRALCSPACSAFPRGTISKQ 147
 Db 170 ARDMLSALYRCDSLVSFQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGLV 229
 Qy 148 D--VCKTCLTRQPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLRVHLVRDP 205
 Db 230 DDRVCKK-CPQRLARFEBCRYRTLVIKGVVFDVAVLAPLLRDPALDLKVIHLVRDP 288
 Qy 206 RAVLRS-----REAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
 Db 289 RAVASSIRSRHGLRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGVGPGADYHALGA 348
 Qy 243 IREVCVSHVRIAEATLKPPFLGRVRLVRFPEDLAREPLAEIRALYAF 302
 Db 349 MEVICNSMAKTLQTA-LQPPDMWQGHYLVVRYEDLVGDPVKTLLRYDYFVGLLVSPMEQ 407
 Qy 303 WIHNITHGSG-IGKPIEAHTSSRNARNVSOAQRHALPFTKILRVOEVCAGALQLLGYR 361
 Db 408 FALNMTSGSSSKP---FVVSARNATQANAWRTALTFOQIKQVEEFCYQPMVAVLYG 464
 Qy 362 VYSADQORDTLDLV 376
 Db 465 VNSPEVKLSKTL 479

RESULT 11
 US-09-471-867-4
 ; Sequence 4, Application US/09471867
 ; Patent No. 6455289
 ; GENERAL INFORMATION:
 ; APPLICANT: Uchimura, Kenji
 ; APPLICANT: Muramatsu, Hideki
 ; APPLICANT: Kadomatsu, Kenji
 ; APPLICANT: Kannagi, Reiji
 ; APPLICANT: Habuchi, Osami
 ; APPLICANT: Muramatsu, Takashi
 ; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
 ; FILE REFERENCE: TOYAMA1.001AUS
 ; CURRENT APPLICATION NUMBER: US/09/471,867
 ; CURRENT FILING DATE: 1999-12-23
 ; EARLIER FILING DATE: 1999-03-05
 ; EARLIER FILING DATE: 1999-03-05
 ; EARLIER FILING DATE: 1998-03-05
 ; EARLIER FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 484
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-471-867-4

Query Match 32.0%; Score 656.5; DB 4; Length 484;
 Best Local Similarity 39.7%; Pred. No. 2.3e-63;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

Qy 34 SPAGGEDRVH-VLVLSWSRSGSFLQOLFSDHFDVFLMEPAWHVMTTL 92
 Db 110 APEGVGDKRHMVYVFTWRSRSGSFFGELFNQNPVEVFLYEPVHVMQKLYPGDAVSLOGA 169
 Qy 93 VRDLMSIFLCMDVFDAYMPQ---SRNLS--AFFNWTSRALCSPACSAFPRGTISKQ 147
 Db 170 ARDMLSALYRCDSLVSFQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGLV 229
 Qy 148 D--VCKTCLTRQPFSLAREACRSYSHVVLKEVRFNQLVLYPLLSDPALNLRVHLVRDP 205

Db 230 DDRVCKK-CPQRLARFECECKRYTLVIGKVRVDFDVAVLAPLRLDPALDLKVIHLVRDP 288
 QY 206 RAVLRS-----REAAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
 Db 289 RAVASSIRSRHGILRESLQVRSRDRPRAHMPFLEAAGHKLGAKKEGVGGPDYHALGA 348
 QY 243 IREYCRSHVRIAEATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFTGLTLTPOLEA 302
 Db 349 MEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVDFVGLLVSPMEQ 407
 QY 303 WIHNIHSG-IGKPIEAFHTSSNARNVQAMRHALPFTKILRVQEVCGAGALLGYRP 361
 Db 408 FALNMTSGSSSKP---FVVSARNATQANAWRTALTFOQIKQVEFCYQPMVAVLGYER 464
 QY 362 VYSADQQRDLTLDIV 376
 Db 465 VNSPEEVKDKSKTL 479

RESULT 12
 US-09-949-016-6471
 ; Sequence 6471, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CL001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6471
 ; LENGTH: 531
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-949-016-6471

Query Match 32.0%; Score 656.5; DB 4; Length 531;
 Best Local Similarity 35.7%; Pred. No. 2.7e-63;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGEDRVH-VLVSSWRSGSFLGQFSDHPDVFYLMPEPAWHVMTTLQSGSAATLHMA 92
 Db 157 APEGVGDKRWVYFTTWRSGSFFGLFNQNPVFFLYEPVHVWQKLYPGDAVSQGA 216
 QY 93 VRDLMSRIFLCMDVFDAYMPQ----SRNLS--AFFNWSRALCSPACSAFPRGTISKQ 147
 Db 217 ARDMSALRYCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKWCSSPLCPAYRKEVVGLV 276
 QY 148 D--VCKTLCTROPFSLAREACRSYSHVVLKVEFRFNQVLYPLISDPALNLRIVHLVRDP 205
 Db 277 DDRVCKK-CPQRLARFECECKRYTLVIGKVRVDFDVAVLAPLRLDPALDLKVIHLVRDP 335
 QY 206 RAVLRS-----REAAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
 Db 336 RAVASSIRSRHGILRESLQVRSRDRPRAHMPFLEAAGHKLGAKKEGVGGPDYHALGA 395
 QY 243 IREYCRSHVRIAEATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFTGLTLTPOLEA 302
 Db 396 MEVICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRVDFVGLLVSPMEQ 454
 QY 303 WIHNIHSG-IGKPIEAFHTSSNARNVQAMRHALPFTKILRVQEVCGAGALLGYRP 361
 Db 455 FALNMTSGSSSKP---FVVSARNATQANAWRTALTFOQIKQVEFCYQPMVAVLGYER 511
 QY 362 VYSADQQRDLTLDIV 376

Db 512 VNSPEEVKDKSKTL 526

RESULT 13
 US-08-899-514-2
 ; Sequence 2, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, MASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; TITLE OF INVENTION: FOR THE POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
 ; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
 ; CITY: NEWPORT BEACH
 ; STATE: CALIFORNIA
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/899,514
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DANIEL E ALTMAN
 ; REGISTRATION NUMBER: 34,115
 ; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 714 760 0404
 ; TELEFAX: 714 760 9502
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 479
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-899-514-2

Query Match 28.2%; Score 577.5; DB 2; Length 479;
 Best Local Similarity 35.7%; Pred. No. 1.2e-54;
 Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

QY 32 PSSPAGGEDRVHVLVSSWRSGSFLGQFSDHPDVFYLMPEPAWHVMTTL-----QGSAA 87
 Db 123 PPRPAVAGPRRHVLLMATTRTGSSFVGFEFNQGNIFLYFLPELWHIERTVSPGPGANAA 182
 QY 88 TLHMAVRDLMSRIFLCMDVFDAY---MQSRNLSAFFNWSRALCSPACSAFPRGTI 144
 Db 183 GSALVYRDVLKQLFLCDLYLLEHFTPLPEDHLTQFMFRGSSRSCLCEDPVCTPFVKKVF 242
 QY 145 SKQDVCKTLCTROPFSLAREACRSYSHVVLKVEFRFNQVLYPLISDPALNLRIVHLVRD 204
 Db 243 EKHCKNRKCGPLNVTLAAEACRKEHMAKAVIRQLQFLQFLAEDPRDLRLVQLVRD 302
 QY 205 PRAVLRSREAAGPILARDNGIVLG---TNGKWVE-----ADPHLRLIREVCRSHVRIA 254
 Db 303 PRAVLASRMA-----PAGKYTKWKWLDDEGQDGLREEVQRLAGNCES-IRLS 351
 QY 255 EAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFTGLTLTPOLEAHTHNTTH-----G 310
 Db 352 AELGLRQPAWLGRVYLVRYEDVARGFLQKAREMYPFAGIPLTPQVEDWTQKNTQAAHDG 411

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.8883 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-7

Perfect score: 2057

Sequence: 1 MRLPRFSSTVMSLLMVQTG.....LPRGMSFKWASSTKQPES 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1708	83.0	395	4	US-09-949-016-7011
2	1536.5	74.7	390	4	US-09-949-016-6813
3	1536.5	74.7	431	4	US-09-949-016-8893
4	1017.5	49.5	386	3	US-09-045-284A-2
5	1017.5	49.5	386	3	US-09-180-911-1
6	986.5	48.0	386	4	US-09-786-240-11
7	647.5	31.5	484	3	US-09-263-023-4
8	647.5	31.5	484	4	US-09-471-867-4
9	647.5	31.5	531	4	US-09-949-016-6471
10	640	31.1	483	3	US-09-263-023-2
11	640	31.1	483	4	US-09-471-867-2
12	636.5	30.9	608	4	US-09-949-016-9449
13	561.5	27.3	479	2	US-08-899-514-2
14	549	26.7	411	3	US-09-015-188-2
15	495	24.1	458	2	US-08-655-878-2
16	96	4.7	468	4	US-09-489-039A-14323
17	91	4.4	487	4	US-09-252-931A-23907
18	90	4.4	933	1	US-08-370-193A-8
19	90	4.4	933	3	US-09-271-438A-9
20	90	4.4	933	4	US-10-078-107-5
21	90	4.4	933	4	US-10-077-751-5
22	89.5	4.4	294	4	US-09-489-039A-12157
23	89.5	4.4	471	4	US-09-252-931A-18083
24	89	4.3	421	4	US-09-902-540-10244
25	88.5	4.3	268	4	US-09-818-780-17
26	88.5	4.3	268	4	US-09-818-780-94
27	88.5	4.3	561	4	US-09-252-931A-24244

28	88.5	4.3	1395	2	US-08-687-399-7	Sequence 7, Appli
29	88	4.3	675	4	US-09-902-540-13407	Sequence 13407, A
30	87.5	4.3	575	4	US-09-252-991A-32986	Sequence 32986, A
31	87.5	4.3	1098	3	US-08-946-994-17	Sequence 17, Appli
32	87	4.2	984	2	US-08-673-789-9	Sequence 9, Appli
33	85	4.1	938	4	US-09-489-039A-13504	Sequence 13504, A
34	85	4.1	1299	5	PCT-US95-08354A-2	Sequence 2, Appli
35	84.5	4.1	1338	4	US-09-631-603-2	Sequence 2, Appli
36	84	4.1	496	3	US-09-292-768-64	Sequence 64, Appli
37	84	4.1	984	2	US-08-449-645A-19	Sequence 19, Appli
38	84	4.1	984	2	US-08-702-367A-19	Sequence 19, Appli
39	84	4.1	984	5	PCT-US95-04681-19	Sequence 2, Appli
40	84	4.1	1248	4	US-10-042-810-2	Sequence 2, Appli
41	84	4.1	1278	4	US-10-042-810-4	Sequence 4, Appli
42	83.5	4.1	460	4	US-09-949-016-7994	Sequence 7994, Ap
43	83.5	4.1	671	4	US-09-248-796A-17595	Sequence 17595, A
44	83.5	4.1	2867	4	US-09-902-540-12593	Sequence 12593, A
45	83	4.0	474	4	US-09-248-796A-17534	Sequence 17534, A

ALIGNMENTS

RESULT 1
US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7011

Query Match 83.0%; Score 1708; DB 4; Length 395;
Best Local Similarity 83.0%; Pred. No. 1.4e-181;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy	1	MRLPRFSSTVMSLLMVQTGILVLSRVQVPSPPAGLGERVHVVLSSWSRSGSFVGOLF	60
Db	1	MMLPRVSTAVTALLAQTELLFLVSRPGSPAGGEARVHVVLSSWSRSGSFVGOLF	60
Qy	61	SQPDVFLMEPAHWHVMDTLCQSAPALHMAVRDLIRSVFLCDMDVDAYLPMRNTSDL	120
Db	61	NQHPDVFYLMPEAHWHVMDTLCQSAAHLHMAVRDLVRSVFLCDMDVDAYLPMRNTSDL	120
Qy	121	QWAVSRALCSPVCEAFARGNISSEEVCKPLCATRPFGLAQACSSYSHVVLKEVRFN	180
Db	121	QWAVSRALCSPACSAFPRGAISSSEAVCKPLCAQSFLLAREACRSYSHVVLKEVRFN	180
Qy	181	LQVLYPLSDPALNRIHVLRDPRVAVLRREQTAKALARDNGIVLTNGTWEADPLR	240
Db	181	LQVLYPLSDPALNRIHVLRDPRVAVLRREQTAKALARDNGIVLTNGTWEADPLR	240
Qy	241	VVNEVCESHVRIAEALHKPPFLQDRVLRVRYEDLARDPLTVIRELYAFTGLTLPOLQ	300
Db	241	VVNEVCESHVRIAEALHKPPFLQDRVLRVRYEDLAREPLAEIRLYAFTGLTLPOLQ	300
Qy	301	TIWNTHGSGPGARREAFKTTSRDALSVSQAWRHHTLPFAKIRVQELCGGALQLGYRS	360

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 1.9e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MRLPFSSTVWLSLMMVQTG-ILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFGQL 59
DB 1 MWLPFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVVLSSWRSGSSFGQL 60
QY 60 FSQHPDVFYLMPEPAWHVWTTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYMPQSRNLSA 120
QY 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPPGLAQEACSSYSHVVVLKEVRFF 179
DB 121 FFWNATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVLKEVRFF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREQTAKALARDNGIVLGTNGTWVEADPRL 239
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREAGPILARDNGIVLGTNGKWEADPHL 240
QY 240 RVNVECRSHVRIAEAAHLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLTPQL 300
QY 300 QTWHNITHGSGPGARREAFKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGYR 359
DB 301 EAWIHNITHSGGIGKPIEAFTSSRNARNVQAWRHLPFTKILRVQEVCAAGALQLLGYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSFKWS 387
DB 361 PVSADQQRDLTLDLVLPGRDPHFSWAS 388

RESULT 3
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match 74.7%; Score 1536.5; DB 4; Length 390;
Best Local Similarity 76.0%; Pred. No. 1.9e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MRLPFSSTVWLSLMMVQTG-ILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFGQL 59
DB 1 MWLPFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVVLSSWRSGSSFGQL 60
QY 60 FSQHPDVFYLMPEPAWHVWTTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYMPQSRNLSA 120
QY 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPPGLAQEACSSYSHVVVLKEVRFF 179
DB 121 FFWNATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVLKEVRFF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREQTAKALARDNGIVLGTNGTWVEADPRL 239
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREAGPILARDNGIVLGTNGKWEADPHL 240
QY 240 RVNVECRSHVRIAEAAHLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 241 RLIREVCRSHVRIAEATLKPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLTPQL 300
QY 300 QTWHNITHGSGPGARREAFKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGYR 359
DB 301 EAWIHNITHSGGIGKPIEAFTSSRNARNVQAWRHLPFTKILRVQEVCAAGALQLLGYR 360
QY 360 SVHSELEQRDLSDLLPRGMDSFKWS 387
DB 361 PVSADQQRDLTLDLVLPGRDPHFSWAS 388

RESULT 3
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8893

Query Match 74.7%; Score 1536.5; DB 4; Length 431;
Best Local Similarity 76.0%; Pred. No. 2.2e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MRLPFSSTVWLSLMMVQTG-ILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFGQL 59
DB 42 MWLPFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVVLSSWRSGSSFGQL 101
QY 60 FSQHPDVFYLMPEPAWHVWTTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
DB 102 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYMPQSRNLSA 161
QY 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPPGLAQEACSSYSHVVVLKEVRFF 179
DB 162 FFWNATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVLKEVRFF 221
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREQTAKALARDNGIVLGTNGTWVEADPRL 239
DB 222 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREAGPILARDNGIVLGTNGKWEADPHL 281
QY 240 RVNVECRSHVRIAEAAHLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 282 RLIREVCRSHVRIAEATLKPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLTPQL 341
QY 300 QTWHNITHGSGPGARREAFKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGYR 359
DB 342 EAWIHNITHSGGIGKPIEAFTSSRNARNVQAWRHLPFTKILRVQEVCAAGALQLLGYR 401
QY 360 SVHSELEQRDLSDLLPRGMDSFKWS 387
DB 402 PVSADQQRDLTLDLVLPGRDPHFSWAS 429

RESULT 4
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.5%; Score 1017.5; DB 3; Length 386;
Best Local Similarity 56.4%; Pred. No. 1.5e-104;
```

```

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8893
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8893

Query Match 74.7%; Score 1536.5; DB 4; Length 431;
Best Local Similarity 76.0%; Pred. No. 2.2e-162;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MRLPFSSTVWLSLMMVQTG-ILVFLVSRQVPSPAGLGERVHVVLSSWRSGSSFGQL 59
DB 42 MWLPFSSKTVTVLLLAQTCLLLFIISRPSPAGGEDRVHVVLSSWRSGSSFGQL 101
QY 60 FSQHPDVFYLMPEPAWHVWTTLSQGSAPALHMAVRDLIRSVFLCDMDVDFDAYLPWRRNISD 119
DB 102 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLIRSVFLCDMDVDFDAYMPQSRNLSA 161
QY 120 LFQWAVSRALCSPPVCEAFARGNISSEVCKPLCATRPPGLAQEACSSYSHVVVLKEVRFF 179
DB 162 FFWNATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVVLKEVRFF 221
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREQTAKALARDNGIVLGTNGTWVEADPRL 239
DB 222 NLQVLYPLSDPALNLRIVHLVRDPAVLRBREAGPILARDNGIVLGTNGKWEADPHL 281
QY 240 RVNVECRSHVRIAEAAHLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLGLTPQL 299
DB 282 RLIREVCRSHVRIAEATLKPPLRGYRLVRFEDLAREPLAEIRALYAFGLTLTPQL 341
QY 300 QTWHNITHGSGPGARREAFKTTSRDALSVQAWRHLPFAKIRRVQELCGGALQLLGYR 359
DB 342 EAWIHNITHSGGIGKPIEAFTSSRNARNVQAWRHLPFTKILRVQEVCAAGALQLLGYR 401
QY 360 SVHSELEQRDLSDLLPRGMDSFKWS 387
DB 402 PVSADQQRDLTLDLVLPGRDPHFSWAS 429

RESULT 4
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107US1
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.5%; Score 1017.5; DB 3; Length 386;
Best Local Similarity 56.4%; Pred. No. 1.5e-104;
```

Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSSTVMSLLMVQTGILVF---LVSRQVPS-SPAGLGERVHVHVLVLSWSSGSSVF 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSKKAQPERMHVHVLVLSWSSGSSVF 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYLMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQQPFVEVKEKACRSYSHVVLKE 178
 Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV-E 234
 Db 179 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV-E 238
 Qy 235 ADPRLRVNVEVCRSHVRIAEAAHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 297
 Qy 295 LTPOLQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGALQ 354
 Db 298 FLPHLQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGALQ 355
 Qy 355 LLGYRSVHSELEORDLSLDLL 375
 Db 356 LLGYRHVRSEQORNLDDLL 376

RESULT 5

US-09-190-911-1
 ; Sequence 1, Application US/09190911
 ; Patent No. 6365365
 ; GENERAL INFORMATION:
 ; APPLICANT: Bistrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kireten
 ; APPLICANT: Hemmerlich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/09/190,911
 ; EARLIER FILING DATE: 1998-11-12
 ; EARLIER APPLICATION NUMBER: 09/045,284
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: H. sapiens
 US-09-190-911-1

Query Match 49.5%; Score 1017.5; DB 3; Length 386;
 Best Local Similarity 56.4%; Pred. No. 1.5e-104;
 Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

Qy 1 MRLPRFSSTVMSLLMVQTGILVF---LVSRQVPS-SPAGLGERVHVHVLVLSWSSGSSVF 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSKKAQPERMHVHVLVLSWSSGSSVF 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYLMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKE 175
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDEIIPRAHCRLLCSQQPFVEVKEKACRSYSHVVLKE 178
 Qy 176 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV-E 234
 Db 179 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV-E 238

Qy 235 ADPRLRVNVEVCRSHVRIAEAAHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGLG 294
 Db 239 EDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 297
 Qy 295 LTPOLQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGALQ 354
 Db 298 FLPHLQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGALQ 355
 Qy 355 LLGYRSVHSELEORDLSLDLL 375
 Db 356 LLGYRHVRSEQORNLDDLL 376

RESULT 6

US-09-786-240-11
 ; Sequence 11, Application US/09786240
 ; Patent No. 6558935
 ; GENERAL INFORMATION:
 ; APPLICANT: INCYTE PHARMACEUTICALS, INC.
 ; APPLICANT: TANG, Y. Tom
 ; APPLICANT: CORLEY, Neil C.
 ; APPLICANT: GUEGLER, Karl J.
 ; APPLICANT: BAUGHN, Mariah R.
 ; APPLICANT: LAL, Preeti
 ; APPLICANT: YUE, Henry
 ; APPLICANT: HILLMAN, Jennifer L.
 ; APPLICANT: AZIMZAI, Valda
 ; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
 ; FILE REFERENCE: PF-0592 PCT
 ; CURRENT APPLICATION NUMBER: US/09/786,240
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
 ; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 11
 ; LENGTH: 386
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
 US-09-786-240-11

Query Match 48.0%; Score 986.5; DB 4; Length 386;
 Best Local Similarity 56.3%; Pred. No. 4.2e-101;
 Matches 215; Conservative 40; Mismatches 114; Indels 13; Gaps 9;

Qy 1 MRLPRFSSTVMSLLMVQTGILVF---LVSRQVPS-SPAGLGERVHVHVLVLSWSSGSSVF 56
 Db 1 MLLPK--KMKLLFLVVSQMAILALFFHMYSHNISLSKKAQPERMHVHVLVLSWSSGSSVF 58
 Qy 57 GOLFSQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYL-PWRR 115
 Db 59 GOLFGQHPDVFYLMPEPAWHVMDTSLQSGAPALHMAVRDLIRSVFLCDMDVFDAYLMEPGPR 118
 Qy 116 NISDLFQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLK 174
 Db 119 RQSSLFQWNSRALCSAPACDIIPQDE-SSFGLTAGSCAVNSPLKLEKACRSYSHVVLK 177
 Qy 175 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV- 233
 Db 178 VRFFNLQVLYPLSDPALNLRIVHLVRDPRAVLSREOTAKALARDNGIVLGTNGTWV- 237
 Qy 234 BADPRLRVNVEVCRSHVRIAEAAHKKPPFPLODRYLRVRYEDLARDPLTVIRELYAFTGL 293
 Db 238 KEDQPYVMQVICSQLEIYK-TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGL 296
 Qy 294 GLTQLOQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGAL 353
 Db 297 EFLPHLQVTHNIHITGSGPGARREAFKTTSDALSVSQAWRHTLPPAKIRRVQELCGGAL 354


```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6471

Query Match      31.1%; Score 647.5; DB 4; Length 531;
Best Local Similarity 36.8%; Pred. No. 4.7e-63;
Matches 140; Conservative 76; Mismatches 117; Indels 47; Gaps 9;

Qy 33 SPAGLGRVH-VLVLSWRGSSVFGOLFQSHQDPVFLMEPAWHVDTLSQGSAPALHMA 91
Db 157 APEGVGDKRHWYVFTTWRSGSPFGELEFNQNPVEFFLYEPVHVMQKLYPGDAVSLQGA 216

Qy 92 VRDLIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NIS 144
Db 217 ARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATNKVWCSSPLCPAYRKEVVGLV 276

Qy 145 SEEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFNFQVLPPLSDPALNLRIVHLVRDP 204
Db 277 DDRVCKK-CPQRLARFECECKRYRTLVKGVRFVDAVLAPLRDPALDLKVIHLVRDP 335

Qy 205 RAVLRSEQTAKALARDNGIVLGTNGTWVEADPR----- 238
Db 336 RAVASSRIRSHGLIRSLQVRSR-----DPAHRMPFLEAAGHKLGAKEGVGGPAD 389

Qy 239 ---LRVNEVCRSHVRIAEALHKPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLGL 295
Db 390 YHALGMEVICNSMAKTLQTL-OPPDWLOQHLYLVRYEDLVGDPVKTLRRVYDFVGLLV 448

Qy 296 TPQLOTWIHNTHGSGPGARREAFKTSRDALSVSQAWRHTLPFAKIRRVQELCGGALQL 355
Db 449 SPEMEQFALNMTSGSGSSK--PFVVSARNATQAAANAWRTALTFOQIKQVEEFCYQPMV 506

Qy 356 LGYRSVHSELQRDLSDLL 375
Db 507 LGYRVNSPPEVKDLSKTL 526

RESULT 10
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

Query Match      31.1%; Score 640; DB 3; Length 483;
Best Local Similarity 36.2%; Pred. No. 2.8e-62;
Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

Qy 30 VPSSPAG-----LG-----ERVHVLVLSWRGSSVFGOLFQSHQDPVFLME 71
Db 89 VPAAGSAGAAASLGNATRGTRGGDKRQLVYVFTTWRSGSPFGELEFNQNPVEFFLYE 148

Qy 72 PAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVS 126
Db 149 PVHVMQKLYPGDAVSLQGAARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATN 208

Qy 127 RALCSPVCEAFARG--NISSEEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFNFQV 184
Db 157 APEGVGDKRHWYVFTTWRSGSPFGELEFNQNPVEFFLYEPVHVMQKLYPGDAVSLQGA 216

Qy 185 YPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPR----- 238
Db 268 APLKDPALDLKVIHLVRDPRAVASSRIRSHGLIRSLQVRSR-----DPAHRMPF 321

Qy 239 ---LRVNEVCRSHVRIAEALHKPPFLQDRYRLVRYED 275
Db 322 LEAAGHKLGAKEGMPGADYHALGMEVICNSMAKTLQTL-OPPDWLOQHLYLVRYED 380

Qy 276 LARDPLTVIRELYAFTGLGLTPQLOTWIHNTHGSGPGARREAFKTSRDALSVSQAWRH 335
Db 381 LVGDPVKTLRRVYDFVGLLVSPEMEQLFALNMTSGSGSSK--PFVVSARNATQAAANAWRT 438

Qy 336 TLPFAKIRRVQELCGGALQLLGYRSVHSELQRDLSDLL 375
Db 439 ALTFQIKQVEEFCYQPMVAVLGYRVNSPPEVKDLSKTL 478

RESULT 11
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match      31.1%; Score 640; DB 4; Length 483;
Best Local Similarity 36.2%; Pred. No. 2.8e-62;
Matches 145; Conservative 73; Mismatches 118; Indels 64; Gaps 10;

Qy 30 VPSSPAG-----LG-----ERVHVLVLSWRGSSVFGOLFQSHQDPVFLME 71
Db 89 VPAAGSAGAAASLGNATRGTRGGDKRQLVYVFTTWRSGSPFGELEFNQNPVEFFLYE 148

Qy 72 PAHWVDTLSQGSAPALHMAVRDLIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVS 126
Db 149 PVHVMQKLYPGDAVSLQGAARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATN 208

Qy 127 RALCSPVCEAFARG--NISSEEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFNFQV 184
Db 157 APEGVGDKRHWYVFTTWRSGSPFGELEFNQNPVEFFLYEPVHVMQKLYPGDAVSLQGA 216

Qy 185 YPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPR----- 238
Db 268 APLKDPALDLKVIHLVRDPRAVASSRIRSHGLIRSLQVRSR-----DPAHRMPF 321

Qy 239 ---LRVNEVCRSHVRIAEALHKPPFLQDRYRLVRYED 275
Db 322 LEAAGHKLGAKEGMPGADYHALGMEVICNSMAKTLQTL-OPPDWLOQHLYLVRYED 380

Qy 276 LARDPLTVIRELYAFTGLGLTPQLOTWIHNTHGSGPGARREAFKTSRDALSVSQAWRH 335
Db 381 LVGDPVKTLRRVYDFVGLLVSPEMEQLFALNMTSGSGSSK--PFVVSARNATQAAANAWRT 438

Qy 336 TLPFAKIRRVQELCGGALQLLGYRSVHSELQRDLSDLL 375
Db 439 ALTFQIKQVEEFCYQPMVAVLGYRVNSPPEVKDLSKTL 478

```

Db 209 KVCSSPLCPAYRKEVVGLVDVDRCKK-CPPQRLARPEEBCRYRTVVIKGVRFVDFVAVL 267
 Qy 185 YPLSDPALNLRIVHLVRDRAVLRSQEQAKALARDNGIVLTNGTWTWEADPR----- 238
 Db 268 APLKOPALDOLKVIHLVRDRAVASSIRSRKHGILRESLQVRSR-----DPAHRMPF 321
 Qy 239 -----LRVNEVCRSHVRIAEEAALHKPPPPFLQDRYRLVRYED 275
 Db 322 LEAAGHKLGAKEGCGPADYHALGAMEVICNSMAKTLQAL-QPPDMLQGHYLVRYED 360
 Qy 276 LARDPLTVIRELYAFTGLTGPQLOTWIHNITGSGPGARREAFKTTSDALSVSQAWRH 335
 Db 381 LVGDPVKTLRRVYDFVGLLVSPMEQPALNMTSGSGSSK--PFWVSARNATQAAWART 438
 Qy 336 TLPKAKIRRYQELCGGALQLLGLVRSVHSELEQRDLSLDLL 375
 Db 439 ALTFQKIQVEEFCYQPMALVGLVRSVNSPEEVKDLSTLL 478

RESULT 12
 US-09-949-016-9449
 ; Sequence 9449, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: CU001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 9449
 ; LENGTH: 608
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-949-016-9449

Query Match 30.9%; Score 636.5; DB 4; Length 608;
 Best Local Similarity 37.1%; Pred. No. 9.8e-62;
 Matches 140; Conservative 73; Mismatches 117; Indels 47; Gaps 9;

Qy 36 GLGE-RVHVLVLSWRSGSFVQGFQHPDVFYLMPEPAWHVMDTLSSQGSAPALHMAVRD 94
 Db 237 GVGDKRQLVVFTWRSSSFFGELFNQNPVEFVYEPVHVWQKLYPGDAVSLOGNARD 236
 Qy 95 LIRSVFLCDMDVFDAYLP---WRRNISDL--FQWAVSRALCSPVCEAFARG--NTISSE 147
 Db 297 MLSALYRCDSLVSFOLYSPAGSGGRNLTGIFGAATNKVVCSSPLCPAYRKEVVGLVDDR 356
 Qy 148 VCKPLCATRPFGLAQEACSSVSHVVLKEVFFNLQVLYPLSDPALNLRIVHLVRDPRV 207
 Db 357 VCKK-CPQRLARPEEBCRYRTLVIKGVRFVDFVAVLAPLRLDPALDVKIHLVRDPRV 415
 Qy 208 LRSEQAKALARDNGIVLTNGTWTWEADPR----- 238
 Db 416 ASSRIRSRHGLRESLQVRSR-----DPAHRMPFLEAGHKLGNKKEGCVGPADYHA 469
 Qy 239 LRVNVEVCRSHVRIAEEAALHKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLGTPQ 298
 Db 470 LGAMEVICNSMAKTLQAL-QPPDMLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSP 528
 Qy 299 LQTWIHNITGSGPGARREAFKTTSDALSVSQAWRHTLPKAKIRRYQELCGGALQLLYG 358
 Db 529 MEQPALNMTSGSGSSK--PFWVSARNATQAAWARTALTFTQKIQVEEFCYQPMALVIG 586
 Qy 359 RSVHSELEQRDLSLDLL 375

Db 587 ERVNSPEEVKDLSTLL 603
 RESULT 13
 US-08-899-514-2
 ; Sequence 2, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HARUCHI, OSAMI
 ; APPLICANT: FUKUTA, MASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; TITLE OF INVENTION: FOR THE POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
 ; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
 ; CITY: NEWPORT BEACH
 ; STATE: CALIFORNIA
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/899,514
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DANIEL E ALTMAN
 ; REGISTRATION NUMBER: 34,115
 ; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 714 760 0404
 ; TELEFAX: 714 760 9502
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 479
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-899-514-2

Query Match 27.3%; Score 561.5; DB 2; Length 479;
 Best Local Similarity 38.6%; Pred. No. 1.6e-53;
 Matches 142; Conservative 55; Mismatches 136; Indels 35; Gaps 11;

Qy 31 PPSAPGLGRVHVLVLSWRSGSFVQGFQHPDVFYLMPEPAWHVMDTLSSQGSAPAL 88
 Db 123 PPRPAVAGPRRHVLLMATTTRTSSVFGFEFFNQGNIFYLFEPLMHTERTVSPFPGANAA 182
 Qy 89 HMAV--RDLIRSVFLCDMDVFDAY---LPWRRNISDLFQWAVSRALCSPVCEAFARGNI 143
 Db 183 GSALVYRDVLKQLFLCDLYLVEHFIITPLPEDLHTQPMFRGSSRSRLCEDPVCTPFVK-KV 241
 Qy 144 SSEEVCCK----PLCATRPFGLAQEACSSVSHVVLKEVFFNLQVLYPLSDPALNLRIV 198
 Db 242 FEKYHCNRCRCPGLNVT-----LAAEACRRKEHMAKAVIRQLFELOPLAEDPRDLRVI 297
 Qy 199 HLVRDPRVLRSR-----EQAKALARDNGIVLTNGTWTWEADPRLRVNVVCRSHVR 251
 Db 298 QLVRDPRVLRSMWAFAGKYTKWKWLDD-----QDGLREBEVQRLR---GNCSIRL 350
 Qy 252 IAAALHKPPPPFLQDRYRLVRYEDLARDPLTVIRELYAFTGLTGLTQLOTWIHNITGSG 311
 Db 351 SABLGL-RQPAWLRGRVLMRYEDVARGPLQKAREMYFAGIPLTFQVEDWIQNTQAAH 409

Qy 312 PGARREAFKTTSDALSVSQAWHTLPPAKIRRVQELCGGALQLLGYRSVHSELEQRDLS 371
Db 410 DGS---GIYSTQKNSSEQEFKWRFSMPFKLAQVVQAPCGPAMRFLGYKLARDAAALTNRS 466
Qy 372 LDLLLPRG 379
Db 467 VSLLEERG 474

RESULT 14

US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; TITLE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015.188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-015-188-2

Query Match 26.7%; Score 549; DB 3; Length 411;
Best Local Similarity 36.4%; Pred. No. 3e-52;
Matches 136; Conservative 58; Mismatches 160; Indels 20; Gaps 9;

Qy 25 LVSRQVSSPA---GLGERVHVLSSWRSGSSFGQLFSQHPDVFYLMPEAHVMDTL- 80
Db 42 LAERLCESPTFAYNLGRKTHILILATRSQSSFGQLFNQHLDFYLFEPYHVQNTLI 101
Qy 81 ---SQSGAPA---LHMVARDLRSVFLCDMDVFDAYL---PWRNISDLFWQAVSRALC 130
Db 102 PRFTQSGPADRRVYMLGASRDLRLSLDYCDLYFLENYIKPPVNHVTHDRIFRGASVLC 161
Qy 131 SPVCEAFARGNISSEE-VCKPLCATRPFGLAQACSSYSHVLKVEFRFFNLQVLYPLLS 189
Db 162 SRPVDPPGADLVLEEGDCVRKCGLLNLTVAAEACRSHRSHVAKTVRVPVNDRLALVE 221
Qy 190 DPALNLRVHVRPRAVLRSEQTAKALARDNGIVLGTNGTWVADPRLRVNVNVCESH 249
Db 222 DPRLNLKVQLVRDPGRGILASRSETFRDYLRLWRLWYGTGRKPNYLD--VTQLTTVCEDF 279
Qy 250 VRIEAAALHKPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLGLTPQLQTVHNTHG 309
Db 280 SNSYSTGLMR-PPWLKGYKMLVRVEDLARNPKMKTTEEIYFGLPILDSHVARNVQNTRG 338
Qy 310 SGPGARREAFKTTSDALSVSQAWHTLPPAKIRRVQELCGGALQLLGYRSVHSELEQRD 369
Db 339 D-PTLGHKHYCTV-RNSAATAEKWRFRLSYDIVAPQAQACQVLAQLGYKIAASEELKN 396
Qy 370 LSLDLLLLPRGMSDF 383
Db 397 PSVSLVEERDFRP 410

RESULT 15

US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:

STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

Query Match 24.1%; Score 495; DB 2; Length 458;
Best Local Similarity 36.1%; Pred. No. 3.8e-46;
Matches 129; Conservative 59; Mismatches 125; Indels 44; Gaps 13;
Qy 40 RVHVLSSWRSGSSFGQLFSQHPDVFYLMPEAHVMDTL--QGSAPALHMAV--RDL 95
Db 112 RRHVLMMATRTGSSFGVBEFNQGNIFYLFEPLMHIERTVTFEPGGANAVGSAVVRDV 171
Qy 96 IRSVFLCDMDVFDAYL---PWRNISDLFWQAVSRALCSPVC-----EAPARGNISSEE 147
Db 172 LQQLLLCDLYLESFISFSPAPEEHLTAALFRGSSHSLCEEPVCTPSLKKVFEKTHCKNRR 231
Qy 148 VCKPLCATRPFGLAQACSSYSHVLKVEFRFFNLQVLYPLLSDDPALNLRIVHLVRDPRV 207
Db 232 -CGPLNT----LAAEACRRKQHWALKTVIRQLEFLOPLAEDRLDLRIIQLVRDPRV 286
Qy 208 LRSREQTAKALARDNGIVLGTNGTW-----VEADPRLRVNVNEV-----CRSHVRIAEAL 257
Db 287 LVSRMVAFS-----GKYESWKWAAEAGEAPLQ-EDEVQRLRGNCESIRLSAELGL 335
Qy 258 HKPPPLQDRYRLVRVEDLARDPLTVIRELYAFTGLGLTPQLQTVHNTHGSGPGARRE 317
Db 336 -RQPRWLRGYMLVRVEDVARAPLRKALEMYRFAGIHPTQVEEMIRANTQAPQDS---N 391
Qy 318 AFKTTSDALSVSQAWHTLPPAKIRRVQELCGGALQLLGYRSVHSELEQRDLSL 372
Db 392 GIYSTQKNSSEQEFKWRFSIPFKLAQVVQAPCGPAMRFLGYKLASSAQELTNRSLSL 448

Search completed: June 23, 2005, 08:52:10
Job time : 16.8883 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 64.4109 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-9

Perfect score: 3147

Sequence: 1 MNRINGAMVDVDAHYKMF.....LIENICWTLMDRLGYPKFM 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3147	100.0	596	4 AAY72641	Aay72641 Human gly
2	3147	100.0	1222	4 AAY72642	Aay72642 Human gly
3	3142	99.8	1212	6 AAE33541	Aae33541 Human nov
4	3142	99.8	1222	6 ABU11849	Abu11849 Human sec
5	2835	90.1	1207	4 AAY72643	Aay72643 Mouse gly
6	2835	90.1	1207	6 AAE33542	Aae33542 Human nov
7	1821	57.9	755	4 AAB93735	Aab93735 Human pro
8	1821	57.9	755	8 ADP56673	Adp56673 Human pro
9	920	29.2	480	4 AAB94410	Aab94410 Human pro
10	637	20.2	128	4 ADG27796	Adg27796 Human nov
11	615	19.5	314	4 AAB93826	Aab93826 Human pro
12	568	18.0	125	8 ADJ12217	Adj12217 Human sec
13	568	18.0	126	2 AAY14415	Aay14415 Human sec
14	147	4.7	410	7 ADN95522	Adn95522 Human BEC
15	147	4.7	411	2 AAW61100	Aaw61100 Keratan s
16	147	4.7	411	5 AAE25356	Aae25356 Human cho
17	147	4.7	411	6 ABU03503	Abu03503 Angiogene
18	147	4.7	411	7 ADJ68589	Adj68589 Human hea
19	147	4.7	411	8 ADQ18725	Adq18725 Human sof
20	136	4.3	474	3 AAB34722	Aab34722 Human sec
21	136	4.3	625	8 ADL10287	Adl10287 Human pro
22	136	4.3	958	3 AAY51120	Aay51120 Human SAR
23	136	4.3	958	8 ADL81102	Adl81102 Human PRO
24	134	4.3	483	2 AAY31656	Aay31656 Mouse N-a
25	133	4.2	484	2 AAY31657	Aay31657 Human N-a

26	133	4.2	530	4 AAB95367	Aab95367 Human pro
27	133	4.2	530	8 ADQ18590	Adq18590 Human sof
28	133	4.2	531	5 AAU69414	Aau69414 Lung smal
29	133	4.2	531	8 ADL14283	Adl14283 Human NF-
30	128	4.1	183	4 ABB68582	Abb68582 Drosophil
31	126	4.0	388	2 AAY39919	Aay39919 Mouse gly
32	123.5	3.9	385	4 AAY72640	Aay72640 Human gly
33	123.5	3.9	395	5 ABB81554	Abb81554 Human cor
34	123.5	3.9	395	5 AAE15438	Aae15438 Human dru
35	123.5	3.9	395	7 ADI21086	Adi21086 Novel hum
36	122.5	3.9	395	8 ADL61235	Adl61235 Human tyr
37	122.5	3.9	479	7 ABM85237	Abm85237 Human pro
38	122.5	3.9	499	6 ABR41139	Abr41139 Human DIT
39	120.5	3.8	479	2 AAW52863	Aaw52863 Glycosami
40	119.5	3.8	390	4 AAY72639	Aay72639 Human gly
41	119.5	3.8	390	5 ABB81556	Abb81556 Human int
42	119	3.8	481	7 ABM85236	Abm85236 Mouse pro
43	118.5	3.8	418	5 ABB81557	Abb81557 Mouse int
44	118	3.7	395	4 AAY72638	Aay72638 Mouse gly
45	118	3.7	395	5 AAU11275	Aau11275 Murine in

ALIGNMENTS

RESULT 1
AAY72641
ID AAY72641 standard; protein; 596 AA.
XX AC AAY72641;
XX DT 02-MAY-2001 (first entry)
XX DS Human glycosyl sulfotransferase-6 (GST-6) fragment.
XX KW Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection.
OS Homo sapiens.
XX FN WO200106015-A1.
XX PD 25-JAN-2001.
XX PF 19-JUL-2000; 2000WO-US019741.
XX PR 20-JUL-1999; 99US-0144694P.
XX PR 13-JUN-2000; 2000US-00593828.
XX PA (REGC) UNIV CALIFORNIA.
XX XX Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX N-PSDB; AAD02702, AAD02703.
XX FT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications.
XX Claim 3; Fig 3; 128pp; English.
XX CC The present sequence is human glycosyl sulfotransferase-6 (GST-6)
XX fragment. GST is a type 2 membrane protein useful for inhibiting a
XX binding event between a selectin and a selectin ligand, which comprises
XX contacting the selectin with a non-sulphated selectin ligand, GST and a
XX small molecular agent that inhibits the sulphation activity of GST. GST

CC is also useful in inhibiting a selectin mediated binding event. GST is
CC useful in gene therapy to treat disorders such as acute or chronic
CC inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis,
CC polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis,
CC diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome,
CC Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism,
CC pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative
CC colitis, dermatitis, myocarditis, regional enteritis, adult respiratory
CC distress syndrome, infantile eczema, psoriasis lichen planus, allergic
CC rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue
CC rejection during transplantation
XX
XX Sequence 596 AA;

Query Match 100.0%; Score 3147; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.6e-307;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFQVNVTFQMESTIT 60
DB 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFQVNVTFQMESTIT 60
QY 61 RIAYVFGPVYINVSSCRFIDSSNPGLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
DB 61 RIAYVFGPVYINVSSCRFIDSSNPGLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
QY 121 ASVADQGITRFGLTQAIKVPVRHDIRIIFPFQFKENIAGVGLICISLVILTTFQWRFLYS 180
DB 121 ASVADQGITRFGLTQAIKVPVRHDIRIIFPFQFKENIAGVGLICISLVILTTFQWRFLYS 180
QY 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVV 240
DB 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVV 240
QY 241 ITSLPGSGAIELKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEKWSDIRSGHF 300
DB 241 ITSLPGSGAIELKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEKWSDIRSGHF 300
QY 301 RLLRGWQLSVQDTKLQIHILHEPNRGKLAQYFAMNKKQKFKGRRESLPEQRSQMKG 360
DB 301 RLLRGWQLSVQDTKLQIHILHEPNRGKLAQYFAMNKKQKFKGRRESLPEQRSQMKG 360
QY 361 AFDRDABYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
DB 361 AFDRDABYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
QY 421 IYSMLYNSKPSLSYLNKVPHEHLAKLFIEGGKGCNLSNGYAFEYEPRLKELSKSNAV 480
DB 421 IYSMLYNSKPSLSYLNKVPHEHLAKLFIEGGKGCNLSNGYAFEYEPRLKELSKSNAV 480
QY 481 SLISHLWLANTAAALRINTDPTSYQLVKFEDIVHPQKTTERRIFAFGLIPISLASLNQ 540
DB 481 SLISHLWLANTAAALRINTDPTSYQLVKFEDIVHPQKTTERRIFAFGLIPISLASLNQ 540
QY 541 ILFATSNLFLVPEYEGISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 541 ILFATSNLFLVPEYEGISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596

RESULT 2
AAY72642
ID AAY72642 standard; protein; 1222 AA.
XX
AC AAY72642;
XX
DT 02-MAY-2001 (first entry)
XX
DE Human glycosyl sulfotransferase-6 (GST-6).

XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;

KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Domain 851..1223
FT /label= C-terminal_sulfotransferase_domain
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US019741.
XX
XX 20-JUL-1999; 99US-0144694P.
PR 13-JUN-2000; 2000US-00593828.
XX
XX (REGC) UNIV CALIFORNIA.
PA
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI: 2001-138471/14.
XX N-PSDB; AAD02702, AAD02704.
DR
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications.
XX
XX Claim 3; Fig 5B; 128pp; English.
XX
CC The present sequence is human glycosyl sulfotransferase-6 (GST-6). GST is
CC a type 2 membrane protein useful for inhibiting a binding event between a
CC selectin and a selectin ligand, which comprises contacting the selectin
CC with a non-sulphated selectin ligand, GST and a small molecular agent
CC that inhibits the sulphation activity of GST. GST is also useful in
CC inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, distress
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation
XX
SQ Sequence 1222 AA;

Query Match 100.0%; Score 3147; DB 4; Length 1222;
Best Local Similarity 100.0%; Pred. No. 4.9e-307;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFQVNVTFQMESTIT 60
DB 627 MNRNGAMDMVDVDAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFQVNVTFQMESTIT 686
QY 61 RIAYVFGPVYINVSSCRFIDSSNPGLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 120
DB 687 RIAYVFGPVYINVSSCRFIDSSNPGLQISLVNNTHEHVSIVTDYHNLKTRFNYLFGGPF 746
QY 121 ASVADQGITRFGLTQAIKVPVRHDIRIIFPFQFKENIAGVGLICISLVILTTFQWRFLYS 180
DB 747 ASVADQGITRFGLTQAIKVPVRHDIRIIFPFQFKENIAGVGLICISLVILTTFQWRFLYS 806
QY 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVV 240
DB 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVV 866
QY 241 ITSLPGSGAIELKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEKWSDIRSGHF 300

||||| 867 ITSLPGSGABEILKQLFNSSDFLYIRVPTAYIDIPETELEIDSFDVACWKVSDIRSGHF 926
 QY 301 RLLRGWLSQSLVQDTKHLQNLHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 360
 Db 927 RLLRGWLSQSLVQDTKHLQNLHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 986
 QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 420
 Db 987 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 1046
 QY 421 IYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGCNLSNGYAFYEPLRKLSEKSKSNV 480
 Db 1047 IYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGCNLSNGYAFYEPLRKLSEKSKSNV 1106
 QY 481 SLLSHLWLANAALRINTDLPTSQYLVKPFEDIVHPPQKTERIFAFGLGIPLSASLNQ 540
 Db 1107 SLLSHLWLANAALRINTDLPTSQYLVKPFEDIVHPPQKTERIFAFGLGIPLSASLNQ 1166
 QY 541 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRLGYPKFM 596
 Db 1167 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRLGYPKFM 1222

RESULT 3
 ID AAE33541 standard; protein; 1212 AA.
 XX AAE33541;
 AC AAE33541;
 DT 16-APR-2003 (first entry)
 DE Human novel CpG-associated gene 1 (NCAG1) encoded protein #1.
 KW Human; novel CpG-associated gene 1; bipolar disorder; neuroprotective;
 KW NCAG1; mood disorder; chromosome 18.
 XX Homo sapiens.
 XX WO2002101044-A2.
 XX 19-DEC-2002.
 XX 06-JUN-2002; 2002WO-EP006316.
 XX 11-JUN-2001; 2001EP-00202214.
 XX (JANC) JANSSEN PHARM NV.
 XX Del-Favero JPL, Van Broeckhoven C;
 XX WPI; 2003-148807/14.
 XX N-PSDB; AAD50032.
 XX New brain expressed genes (designated novel CpG-associated gene 1
 PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
 PT bipolar or mood disorders, and as targets for developing drugs for the
 PT treatment these disorders.
 XX Claim 14; Col 38-42; 26pp; English.
 XX The invention relates to novel CpG-associated gene 1 (NCAG1) brain-
 CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
 CC encoded protein is useful as a diagnostic marker for bipolar disorder
 CC such as mood disorders. They are also useful as targets for developing
 CC drugs, as well as for target validation, for the treatment of bipolar
 CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
 CC located at chromosome 18
 XX Sequence 1212 AA;

Query Match 99.8%; Score 3142; DB 6; Length 1212;
 Best Local Similarity 99.8%; Pred. No. 1.5e-306;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMESTIT 60
 Db 617 MNRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAEQAABFKKRWTFQVNVVTFQMEPTIT 676
 QY 61 RIAYVFYGPYINVSSCRFDISSNPGLQISLVNNTTEHVSIIVTDYHNLKTRFNYLGRGGF 120
 Db 677 RIAYVFYGPYINVSSCRFDISSNPGLQISLVNNTTEHVSIIVTDYHNLKTRFNYLGRGGF 736
 QY 121 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILICISLVILTFQRFYLS 180
 Db 737 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILICISLVILTFQRFYLS 796
 QY 181 FRKLMRWILILVIALWFIELLDVMSCTSQPICAKWTRTEAGSKSLSSSEGHMDLDPDV 240
 Db 797 FRKLMRWILILVIALWFIELLDVMSCTSQPICAKWTRTEAGSKSLSSSEGHMDLDPDV 856
 QY 241 ITSPLPGSABEILKQLFNSSDFLYIRVPTAYIDIPETELEIDSFDVACWKVSDIRSGHF 300
 Db 857 ITSPLPGSABEILKQLFNSSDFLYIRVPTAYIDIPETELEIDSFDVACWKVSDIRSGHF 916
 QY 301 RLLRGWLSQSLVQDTKHLQNLHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 360
 Db 917 RLLRGWLSQSLVQDTKHLQNLHLHPEPNRGKLAQYFAMNKKRKFKRRESLPEORSQWKG 976
 QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 420
 Db 977 AFDRDAEYIRALRRHLVYPSARPVLSSSGSWTLKLHFFQEVLGASMRALYIYVRDPRAW 1036
 QY 421 IYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGCNLSNGYAFYEPLRKLSEKSKSNV 480
 Db 1037 IYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGCNLSNGYAFYEPLRKLSEKSKSNV 1096
 QY 481 SLLSHLWLANAALRINTDLPTSQYLVKPFEDIVHPPQKTERIFAFGLGIPLSASLNQ 540
 Db 1097 SLLSHLWLANAALRINTDLPTSQYLVKPFEDIVHPPQKTERIFAFGLGIPLSASLNQ 1156
 QY 541 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRLGYPKFM 596
 Db 1157 ILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRLGYPKFM 1212

RESULT 4
 ABU11849
 ID ABU11849 standard; protein; 1222 AA.
 XX ABU11849;
 XX 12-FEB-2003 (first entry)
 XX Human secreted protein SECP-4, INCYTE 3441255CD1.
 DE Human; SECP; secreted protein; micro-array; liver disease; hepatitis;
 KW cirrhosis; cell proliferative disease; cancer; atherosclerosis;
 KW neurological disorder; epilepsy; Huntington's disease; stroke;
 KW cardiovascular disorder; hypertension; angina pectoris; allergy;
 KW myocardial infarction; immune disorder; inflammatory disorder; AIDS;
 KW hypothyroidism; acquired immunodeficiency syndrome; Cushing's syndrome;
 KW developmental disorder.
 XX Homo sapiens.
 XX WO200286069-A2.
 XX 31-OCT-2002.
 XX 19-APR-2002; 2002WO-US012464.
 XX 20-APR-2001; 2001US-0285207P.
 PR 27-APR-2001; 2001US-0287114P.
 PR 03-MAY-2001; 2001US-0288640P.
 PR 11-MAY-2001; 2001US-0290516P.

```

PR 18-MAY-2001; 2001US-0292184P.
PR 21-DEC-2001; 2001US-0343553P.
PR 13-FEB-2002; 2002US-0357002P.
PR 20-FEB-2002; 2002US-0358279P.
PR 19-MAR-2002; 2002US-0366041P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
XX Klammer AA, Hafalia AJA, Duggan BM, Warren BA, Emerling BM;
PI Tribouley CM, Arvizu CS, Honchell CD, Nguyen DB, Kallick DA, Yue H;
PI Au-Young JK, Ramkumar J, Li JX, Thangavelu K, Gbetzen KJ, Ding L;
PI Baughn MR, Yao MG, Walia NK, Mason PM, Lal PG, Graul RC, Reddy R;
PI Becha SD, Sapperstein SK, Richardson TW, Tran UK, Elliott VS;
PI Tang YT, Azimzai Y, Yan L, Xu Y;
XX
XX WPI; 2003-093118/08.
DR N-PSDB; ABX49951.
XX
XX New human secreted proteins (SECP) useful for diagnosing, treating and
PT preventing diseases or conditions associated with the aberrant SECP
PT expression e.g. cancer, AIDS, atherosclerosis, epilepsy, allergies,
PT hepatitis, cirrhosis.
XX
XX Claim 1; Page 137-140; 192pp; English.
XX
XX The invention relates to an isolated polypeptide comprising any of 30
CC secreted human proteins (SECP1-SECP30) appearing as ABU1846-ABU1875, a
CC naturally occurring amino acid sequence at least 90-98 % identical to the
CC sequences, or a biologically active or immunogenic fragment of the
CC polypeptide. Also included are an isolated polynucleotide encoding SECP,
CC (including a polynucleotide sequence at least 90-98 % identical to the
CC sequences, their complements, RNA equivalents or fragments comprising at
CC least 60 contiguous nucleotides) a recombinant polynucleotide comprising
CC a promoter sequence operably linked to the SECP polynucleotide, a cell
CC transformed with the recombinant polynucleotide, a transgenic organism
CC comprising the recombinant polynucleotide, an anti-SECP antibody,
CC screening for ant/agonists of SECP, generating an expression profile of a
CC sample containing the polynucleotides and an array comprising different
CC nucleotide molecules affixed at distinct physical locations on a solid
CC substrate, where at least one nucleotide molecule comprises a first
CC oligonucleotide or polynucleotide sequence specifically hybridisable with
CC at least 30 contiguous nucleotides of the target polynucleotide. The
CC polypeptides and polynucleotides are useful in diagnosing, treating and
CC preventing diseases or conditions associated with the decreased
CC expression or overexpression of SECP, such as liver (e.g. hepatitis,
CC cirrhosis), cell proliferative (e.g. cancer, atherosclerosis),
CC neurological (e.g. epilepsy, Huntington's disease, stroke),
CC cardiovascular (e.g. hypertension, angina pectoris, myocardial
CC infarction), immune/inflammatory (e.g. acquired immunodeficiency syndrome
CC (AIDS), allergies) and developmental (e.g. hypothyroidism, Cushing's
CC syndrome) disorders (many other diseases and conditions are given in the
CC specification). These are also useful in assessing the effects of
CC exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of SECP. The SECP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The micro-array is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles. The
CC present sequence is a SECP protein of the invention
XX
XX Sequence 1222 AA;
XX
XX Query Match 99.8%; Score 3142; DB 6; Length 1222;
XX Best Local Similarity 99.8%; Pred. No. 1.6e-306;
XX Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MNRVNGAMDVWDADHYKWFDFHNGNSPMASIQAEQADEFKRWTFQVNVTFQMSSTIT 60
XX
XX 627 MNRVNGAMDVWDADHYKWFDFHNGNSPMASIQAEQADEFKRWTFQVNVTFQMSPTIT 686
XX
XX 61 RIAYVFYGPYINVSSCRFIDSSNPGLIQISLVNNTTEHVVSIVTDYHNLKTRFNLYGFGGF 120
XX

```

```

Db 687 RIAYVFYGPYINVSSCRFIDSSNPGLIQISLVNNTTEHVVSIVTDYHNLKTRFNLYGFGGF 746
QY 121 ASVADQGQITREGLCTQOAIVKPVVRHDIRIIFPGFKNIAVGLILCISLVILFQWRFYLS 180
Db 747 ASVADQGQITREGLCTQOAIVKPVVRHDIRIIFPGFKNIAVGLILCISLVILFQWRFYLS 806
QY 181 FRKLNRWILILVALWFIELLDVWSTCSQPICAKWTRTRTEAEGSKKLSLSEGHMDLPDVV 240
Db 807 FRKLNRWILILVALWFIELLDVWSTCSQPICAKWTRTRTEAEGSKKLSLSEGHMDLPDVV 866
QY 241 ITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACSWKVSDIRSGHF 300
Db 867 ITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACSWKVSDIRSGHF 926
QY 301 RLIRGLWLSIVQDTKLHLQNIHLHBNRCKLAQYFAMNKDKKPKKRRESLPEQRSOMKG 360
Db 927 RLIRGLWLSIVQDTKLHLQNIHLHBNRCKLAQYFAMNKDKKPKKRRESLPEQRSOMKG 986
QY 361 AFDRDAEYIRALRRHLVYVPSARPVLSLSSGSGWTLLKHFQEVLGASMRALYIVRDPRAW 420
Db 987 AFDRDAEYIRALRRHLVYVPSARPVLSLSSGSGWTLLKHFQEVLGASMRALYIVRDPRAW 1046
QY 421 IYSMLYNSKPSLYSLKQNVPEHLAKLFKIEGKGKCNLNSGYAFEPYELRKLKSKSKSNV 480
Db 1047 IYSMLYNSKPSLYSLKQNVPEHLAKLFKIEGKGKCNLNSGYAFEPYELRKLKSKSKSNV 1106
QY 481 SLLSHLWLTAAARINTDLLPTSQVLVKFEDIYVHFPOKTTTERIFAFILGIPLSASLNQ 540
Db 1107 SLLSHLWLTAAARINTDLLPTSQVLVKFEDIYVHFPOKTTTERIFAFILGIPLSASLNQ 1166
QY 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
Db 1167 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 1222
XX
XX RESULT 5
XX AAY72643
XX ID AAY72643 standard; protein; 1207 AA.
XX AC AAY72643;
XX DT 02-MAY-2001 (first entry)
XX DE Mouse glycosyl sulfotransferase-6 (GST-6).
XX KW Mouse; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;
XX selectin binding inhibitor; gene therapy; inflammation;
XX KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection.
XX OS Mus sp.
XX PN WO200106015-A1.
XX PD 25-JAN-2001.
XX PF 19-JUL-2000; 2000WO-US019741.
XX PR 20-JUL-1999; 99US-0144694P.
XX PR 13-JUN-2000; 2000US-00593828.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX DR N-PSDB; AAD02705, AAD02706.
XX

```


PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications.
 PS Example 2; Fig 6B; 128pp; English.

XX The present sequence is mouse glycosyl sulfotransferase-6 (GST-6). GST is
 CC a type 2 membrane protein useful for inhibiting a binding event between a
 CC selectin and a selectin ligand, which comprises contacting the selectin
 CC with a non-sulphated selectin ligand, GST and a small molecular agent
 CC that inhibits the sulphation activity of GST. GST is also useful in
 CC inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation

XX Sequence 1207 AA;

Query Match 90.1%; Score 2835; DB 4; Length 1207;
 Best Local Similarity 88.6%; Pred. No. 1.4e-275;
 Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDWDAHYKQWFDHNGSPMASIQAEQAEPKRWTFQVNVTFHMESTIT 60
 Db 612 MNRNGAMVDWDAHYKQWFDHNGSPMASIQAEQAEPKRWTFQVNVTFHMESTIT 671
 Qy 61 RIAYVFGPYNVSSCRFIDSSNPGLOISLVNNTTEHVSVTVTDYHNLKTRFNYLPGGFG 120
 Db 672 RIAYVFGPYNVSSCRFIDSSNPGLOISLVNNTTEHVSVTVTDYHNLKTRFNYLPGGFG 731
 Qy 121 ASVADQGOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 180
 Db 732 ASVANQGOITRFGLGTOEIVNPNVRHDKVNFPPGFKFNIAGVGLIICISILVITFQWRFYLS 791
 Qy 181 FRKLMRWILVIALWFIELLDVWSTCTQPTCAKWTREAGSKSLSEGHMDLPDVV 240
 Db 792 FRKLMRWILVIALWFIELLDVWSTCTQPTCAKWTREAGSKSLSEGHMDLPDVV 851
 Qy 241 ITSPLPGSABTLKQLFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
 Db 852 ITSPLPGSABTLKQLFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 911
 Qy 301 RLLRGWLQSLVQDQTKLHNLHLEPNRGLAQYFAMNKKRKFRRRESLPEQORSQMKG 360
 Db 912 HLLRGWLQSLVQDQTKLHNLHLEPNRGLAQYFAMNKKRKFRRRESLPEQORSQMKG 971
 Qy 361 AFDRDAEYIRALRRHLVYYPSPARVLSLSSGSGWTCLKHFFQEVIGASNRALYIVRDPRAW 420
 Db 972 PFDRDAEYIRALRRHLVYYPSPARVLSLSSGSGWTCLKHFFQEVIGASNRALYIVRDPRAW 1031
 Qy 421 IYSMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLSGYAFEPVLPKELSKSNV 480
 Db 1032 IYSVLYGSKPSLYSLKNVPEHLAKLFKEGKGKCNLSGYAFEPVLPKELSKSNV 1091
 Qy 481 SLLSHLWANTAAALRINTDLLPTSYQLVKPEDEVHPQKTERIFAPLGIPLSPASLNQ 540
 Db 1092 SLLSHLWANTAAALRINTDLLPTSYQLVKPEDEVHPQKTERIFAPLGIPLSPASLNQ 1151
 Qy 541 ILFATSTNLFYLPVEGEISPTNTNWKQNLPRDEIKLIENICWTLMDLRGLYPKFMD 596
 Db 1152 MLFATSTNLFYLPVEGEISPTNTNWKQNLPRDEIKLIENICWTLMDLRGLYPKFMD 1207

RESULT 6
 AAE33542
 ID AAE33542 standard; protein; 1207 AA.
 XX

AC AAE33542;
 XX 16-APR-2003 (first entry)
 DT Human novel CpG-associated gene 1 (NCAG1) encoded protein #2.
 DE
 XX Human, novel CpG-associated gene 1; bipolar disorder; neuroprotective;
 KW NCAG1; mood disorder; chromosome 18.
 XX Homo sapiens.
 OS
 XX WP02002101044-A2.
 PN
 XX 19-DEC-2002.
 PD
 XX 06-JUN-2002; 2002MO-EP006316.
 PF
 XX 11-JUN-2001; 2001EP-00202214.
 PR
 XX (JANC) JANSSEN PHARM NV.
 PA
 XX Del-Pavero JPL, Van Broeckhoven C;
 PI
 XX WPI; 2003-148807/14.
 DR
 XX N-PSDB; AAD50033.
 DT
 XX New brain expressed genes (designated novel CpG-associated Gene 1
 PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
 PT bipolar or mood disorders, and as targets for developing drugs for the
 PT treatment these disorders.
 PT
 XX Claim 16; Col 48-51; 26pp; English.
 PS
 XX The invention relates to novel CpG-associated gene 1 (NCAG1) brain-
 CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
 CC encoded protein is useful as a diagnostic marker for bipolar disorder
 CC such as mood disorders. They are also useful as targets for developing
 CC drugs, as well as for target validation, for the treatment of bipolar
 CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
 CC located at chromosome 18
 CC
 XX SQ Sequence 1207 AA;

Query Match 90.1%; Score 2835; DB 6; Length 1207;
 Best Local Similarity 88.6%; Pred. No. 1.4e-275;
 Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDWDAHYKQWFDHNGSPMASIQAEQAEPKRWTFQVNVTFHMESTIT 60
 Db 612 MNRNGAMVDWDAHYKQWFDHNGSPMASIQAEQAEPKRWTFQVNVTFHMESTIT 671
 Qy 61 RIAYVFGPYNVSSCRFIDSSNPGLOISLVNNTTEHVSVTVTDYHNLKTRFNYLPGGFG 120
 Db 672 RIAYVFGPYNVSSCRFIDSSNPGLOISLVNNTTEHVSVTVTDYHNLKTRFNYLPGGFG 731
 Qy 121 ASVADQGOITRFGLGTOAIVKPVHRDRIIPFGFKFNIAGVGLIICISILVITFQWRFYLS 180
 Db 732 ASVANQGOITRFGLGTOEIVNPNVRHDKVNFPPGFKFNIAGVGLIICISILVITFQWRFYLS 791
 Qy 181 FRKLMRWILVIALWFIELLDVWSTCTQPTCAKWTREAGSKSLSEGHMDLPDVV 240
 Db 792 FRKLMRWILVIALWFIELLDVWSTCTQPTCAKWTREAGSKSLSEGHMDLPDVV 851
 Qy 241 ITSPLPGSABTLKQLFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
 Db 852 ITSPLPGSABTLKQLFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 911
 Qy 301 RLLRGWLQSLVQDQTKLHNLHLEPNRGLAQYFAMNKKRKFRRRESLPEQORSQMKG 360
 Db 912 HLLRGWLQSLVQDQTKLHNLHLEPNRGLAQYFAMNKKRKFRRRESLPEQORSQMKG 971
 Qy 361 AFDRDAEYIRALRRHLVYYPSPARVLSLSSGSGWTCLKHFFQEVIGASNRALYIVRDPRAW 420
 Db 972 PFDRDAEYIRALRRHLVYYPSPARVLSLSSGSGWTCLKHFFQEVIGASNRALYIVRDPRAW 1031
 Qy 421 IYSMLYNSKPSLYSLKNVPEHLAKLFKEGKGKCNLSGYAFEPVLPKELSKSNV 480
 Db 1032 IYSVLYGSKPSLYSLKNVPEHLAKLFKEGKGKCNLSGYAFEPVLPKELSKSNV 1091
 Qy 481 SLLSHLWANTAAALRINTDLLPTSYQLVKPEDEVHPQKTERIFAPLGIPLSPASLNQ 540
 Db 1092 SLLSHLWANTAAALRINTDLLPTSYQLVKPEDEVHPQKTERIFAPLGIPLSPASLNQ 1151
 Qy 541 ILFATSTNLFYLPVEGEISPTNTNWKQNLPRDEIKLIENICWTLMDLRGLYPKFMD 596
 Db 1152 MLFATSTNLFYLPVEGEISPTNTNWKQNLPRDEIKLIENICWTLMDLRGLYPKFMD 1207

Db 972 PFORDAEYIRALRRHLVYYPARPVLSLSSGSWTLKLFHFQVGLTSMRALYIVRDPRAW 1031
QY 421 IYMLYNSKPSLYSLKNVPEHLAKLPKIEGKGKCNLNSGVAPEYBPLRKLKSKSNAV 480
Db 1032 IYVLYGSKPSLYSLKNVPEHLAKLPKIEGKGKCNLSNGVAPEYBPLRKLKSKSNAV 1091
QY 481 SLASHLWLAATAALRINTDLPSTYOLVKFEDIVHFPQKTERIFAFGLIPSPASLNQ 540
Db 1092 SLASHLWLAATAALRINTDLPSTYOLVKFEDIVHFPQKTERIFAFGLIPSPASLNQ 1151
QY 541 ILFATSTNLYPLVEGIBSTNTNVKQNLPRDEIKLIENICWTLMDRIGYKPM 596
Db 1152 MLFATSTNLYPLVEGIBSTNTNVKQNLPRDEIKLIENICWTLMDRIGYKPM 1207

RESULT 7
AAB93735
ID AAB93735 standard; protein; 755 AA.
XX AAB93735;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:13382.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 13382; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent CC oligonucleotides, all of which are used in the exemplification of the CC present invention
XX
SQ Sequence 755 AA;

Query Match 57.9%; Score 1821; DB 4; Length 755;
Best Local Similarity 99.4%; Pred. No. 1.1e-173;
Matches 342; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRNGAMMDVMDAHYKMFWDHGHNSPMASIQEABQAAEFKKRWTFQVNVTFQMESTIT 60
Db 411 MNRNGAMMDVMDAHYKMFWDHGHNSPMASIQEABQAAEFKKRWTFQVNVTFQMESTIT 470
QY 61 RIAYFYGPYINVSRCRFDSSNPGLOISLNVNTEHVSIIVTDYHNLKTRNLYLGGGF 120
Db 471 RIAYFYGPYINVSRCRFDSSNPGLOISLNVNTEHVSIIVTDYHNLKTRNLYLGGGF 530
QY 121 ASVADOGQITRFGLGTOAIVKPVHRDRIIFPFQGFKNIAVGLILCISLVILTFQWRFYLS 180
Db 531 ASVADOGQITRFGLGTOAIVKPVHRDRIIFPFQGFKNIAVGLILCISLVILTFQWRFYLS 590
QY 181 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWRTRTEAGSKKSLSEGHMDLPDVV 240
Db 591 FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWRTRTEAGSKKSLSEGHMDLPDVV 650
QY 241 ITSLSGSGAEILKQLPFNSSDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 300
Db 651 ITSLSGSGAEILKQLPFNSSDFLYIRVPTAYIDIPETELEDISFVDACEWKVSDIRSGHF 710
QY 301 RLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKDKKK 344
Db 711 RLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKDKKK 754

RESULT 8

ADP56673

ID ADP56673 standard; protein; 755 AA.

XX ADP56673;

XX 12-AUG-2004 (first entry)

DT Human protein which is a chondroitin 6 sulphotransferase 3 homologue.

DE cancer detection; large intestine; oesophagus; stomach; lungs; pancreas;
KW liver; kidney; colon; human; chondroitin 6 sulphotransferase 3; CH6T3.

XX Homo sapiens.

XX JP2004147505-A.

XX 27-MAY-2004.

XX 28-OCT-2002; 2002JP-00312927.

XX 28-OCT-2002; 2002JP-00312927.

XX (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.

XX (FJRE) FUJIREBIO KK.

XX (SEKG) SEIKAGAKU KOGYO CO LTD.

XX WPI; 2004-434537/41.

XX N-PSDB; ADP56672.

XX Novel nucleic acid such as DNA, useful as probe for identifying or

PT detecting cancerous tissue of esophagus, stomach, lungs, pancreas, liver,

PT kidney or colon, preferably large intestine tissue.

XX Example 1; SEQ ID NO 2; 24pp; Japanese.

XX The invention relates to a novel nucleic acid having 40-1000 base pairs

CC and comprising a sequence which is complementary to 41-2308 nucleotides
 CC of a fully defined sequence of 3288 base pairs as given in the
 CC specification. The polynucleotide of the invention may be useful for
 CC rapid and highly reliable detection of cancerous tissue derived from the
 CC large intestine, oesophagus, stomach, lungs, pancreas, liver, kidney or
 CC colon. The current sequence is that of the human protein of the invention
 CC which is a chondroitin 6 sulphotransferase 3 (CHST3) homologue.
 XX
 SQ Sequence 755 AA;

Query Match 57.9%; Score 1821; DB 8; Length 755;
 Best Local Similarity 99.4%; Pred. No. 1.1e-173; Mismatches 1; Indels 0; Gaps 0;
 Matches 342; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRVNGAMDMVDAAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFVNVTFQMESTIT 60
 DB 411 MNRVNGAMDMVDAAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFVNVTFQMESTIT 470
 QY 61 RIAYVFYGPYINVSSCRFDIDSSNPGQLISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 120
 DB 471 RIAYVFYGPYINVSSCRFDIDSSNPGQLISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 530
 QY 121 ASVADQGITRFGGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWREYLS 180
 DB 531 ASVADQGITRFGGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTFQWREYLS 590
 QY 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVV 240
 DB 591 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWRTEAGSKSLSSGHHMDLPDVV 650
 QY 241 ITSPLGSGAEILKQFPNSSDFLYIRVPTAVIDIPETELEIDSFVDACEWKSVDIRSGHF 300
 DB 651 ITSPLGSGAEILKQFPNSSDFLYIRVPTAVIDIPETELEIDSFVDACEWKSVDIRSGHF 710
 QY 301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKKK 344
 DB 711 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKKK 754

RESULT 9
 ID AAB94410
 XX AAB94410 standard; protein; 480 AA.
 AC AAB94410;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:14997.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-00116126.
 PF
 XX 29-JUL-1999; 99JP-00248036.
 PR
 XX 27-AUG-1999; 99JP-00300253.
 PR
 XX 11-JAN-2000; 2000JP-00118776.
 PR
 XX 02-MAY-2000; 2000JP-00183767.
 PR
 XX 09-JUN-2000; 2000JP-00241899.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-

PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

PS Claim 8; SEQ ID NO 14997; 2537pp + Sequence Listing; English.

XX
 CC The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX

SQ Sequence 480 AA;

Query Match 29.2%; Score 920; DB 4; Length 480;
 Best Local Similarity 100.0%; Pred. No. 3.9e-83;
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRVNGAMDMVDAAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFVNVTFQMESTIT 60
 DB 308 MNRVNGAMDMVDAAHYKMFDFDHGNSPMASIQAEQAEEFKKRWTFVNVTFQMESTIT 367
 QY 61 RIAYVFYGPYINVSSCRFDIDSSNPGQLISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 120
 DB 368 RIAYVFYGPYINVSSCRFDIDSSNPGQLISLVNNTVHVSIVTDYHNLKTRFNYLGFGGF 427
 QY 121 ASVADQGITRFGGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTF 173
 DB 428 ASVADQGITRFGGLGTOAIVKPVHRDRIIPFGFKFNIAVGLILCISLVILTF 480

RESULT 10
 ADG27796
 ID ADG27796 standard; protein; 128 AA.
 XX
 AC ADG27796;
 XX
 DT 26-FEB-2004 (first entry)
 XX
 DE Human novel protein amino acid sequence SeqID563.
 XX
 KW antiparkinsonian; haemostatic; nootropic; neuroprotective; osteopathic;
 KW anti-HIV; protozoacide; antifungal; immunosuppressive; antirheumatic;
 KW antiarthritic; antidiabetic; antiallergic; antiinflammatory;
 KW anticoagulant; cytostatic; gene therapy; Parkinson's disease;
 KW Alzheimer's disease; thrombocytopaenia; osteoporosis; osteoarthritis;
 KW infection; HIV; Leishmania; malaria; fungal infection;
 KW multiple sclerosis; rheumatoid arthritis;
 KW insulin dependent diabetes mellitus; allergic reaction; food allergy;
 KW insect allergy; allergic rhinitis; haemophilia; cancer; human.
 XX
 OS Homo sapiens.
 XX
 FN WO200179254-A1.
 XX
 XX 25-OCT-2001.

```

XX PF 16-APR-2001; 2001WO-US008655.
XX PR
XX PR 18-APR-2000; 2000US-00552929.
XX PR 22-SEP-2000; 2000US-00668317.
XX PR 24-OCT-2000; 2000US-00695783.
XX PR 01-DEC-2000; 2000US-00728628.
XX PR 26-JAN-2001; 2001US-00770160.
XX PR 13-FEB-2001; 2001US-00783066.
XX PR 22-MAR-2001; 2001US-00816928.
XX PA
XX (HYSE-) HYSEQ INC.
XX XX
XX Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
XX Yang Y, Zhao QA, Goodrich RW, Liu C, Drmanac RT, Ma Y, Wang Z;
XX Wehrman T;
XX WPI; 2001-607699/69.
XX PR
XX Novel polynucleotides and encoded polypeptides (protein factors,
XX including, e.g., cytokines, such as lymphokines, interferons, and
XX circulating soluble factors) useful for treating, e.g., Parkinson's,
XX Alzheimer's, HIV and cancer.
XX PR
XX Claim 20; SEQ ID NO 563; 153pp; English.
XX PS
XX This invention relates to a novel isolated DNA sequence and the mature
XX proteins encoded by them. The invention may be useful in the development
XX of compositions with antiparkinsonian, haemostatic, neurotropic,
XX neuroprotective, osteopathic, anti-HIV, protozoacide, antifungal,
XX immunosuppressive, antiarthritis, antidiabetic,
XX antiallergic, antiinflammatory, anticoagulant or cytostatic activities.
XX In addition, the sequences of the invention may be useful for gene
XX therapy. The invention may be useful for the development of treatments
XX for Parkinson's, Alzheimer's, thrombocytopaenia, osteoporosis,
XX osteoarthritis, infections (including HIV, Leishmania, malaria, and
XX various fungal infection), autoimmune disorders such as multiple
XX sclerosis, rheumatoid arthritis, and insulin dependent diabetes mellitus,
XX allergic reactions and conditions (for example food allergies, insect
XX allergies and allergic rhinitis), coagulation disorders including
XX haemophilia, and cancer. Note: The amino acid sequences given in table 6
XX (SeqID 439-584) may have in frame stop codons or possible
XX insertions/deletions as shown in the table. The sequences allocated Seq
XX IDs 1-438 are not provided (even by reference) in the specification.
XX SQ
XX Sequence 128 AA;
XX
XX Query Match 20.2%; Score 637; DB 4; Length 128;
XX Best Local Similarity 96.1%; Pred. No. 1.7e-55;
XX Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
XX QY 470 KELS KSNVSLSHLWLANTAALRINTDLPTSQVLVKFEDIVHFPQKTTTIPAF 529
XX Db 2 KNYQNPKNVSLSHLWLANTAALRINTDLPTSQVLVKFEDIVHFPQKTTTIPAF 61
XX
XX QY 530 GIPLSASLNLQILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRL 589
XX Db 62 GIPLSASLNLQILFATSTNLFYLPYEGEISPTNTNWKNLPRDEIKLIENICWTLMDRL 121
XX
XX QY 590 GYPKFMD 596
XX Db 122 GYPKFMD 128
XX
XX RESULT 11
XX AAB93826
XX ID AAB93826 standard; protein; 314 AA.
XX AC AAB93826;
XX XX
XX 26-JUN-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:13638.

```

```

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX PD 07-FEB-2001.
XX XX
XX PF 28-JUL-2000; 2000EP-00116126.
XX PR
XX 29-JUL-1999; 99JP-00248036.
XX PR 27-AUG-1999; 99JP-00300253.
XX PR 11-JAN-2000; 2000JP-00118776.
XX PR 02-MAY-2000; 2000JP-00183767.
XX PR 09-JUN-2000; 2000JP-00241899.
XX XX
XX (HELI-) HELIX RES INST.
XX XX
XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX DR
XX PR
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX Claim 8; SEQ ID NO 13638; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX polynucleotides, all of which are used in the exemplification of the
XX present invention
XX SQ
XX Sequence 314 AA;
XX
XX Query Match 19.5%; Score 615; DB 4; Length 314;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 224 KKSLSSEGHMDLPDVVITSLPGSGAEILKQLPFNSDFLYIRVPTAYIDIPETEIDS 283
XX Db 198 KKSLSSEGHMDLPDVVITSLPGSGAEILKQLPFNSDFLYIRVPTAYIDIPETEIDS 257
XX
XX QY 284 FVDACEKWSVDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKD 340
XX Db 258 FVDACEKWSVDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPRGKLAQYFAMNKD 314
XX
XX RESULT 12
XX ADJ12217
XX ID ADJ12217 standard; protein; 125 AA.
XX XX

```

AC ADJ12217;
 DT 20-MAY-2004 (first entry)
 DE Human secreted protein SeqID 71.
 KW human; secreted; cancer; haematopoietic disease; anaemia;
 KW multiple myeloma; reproductive system disorder; prostatitis;
 KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;
 KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;
 KW fetal alcohol syndrome; Down's syndrome; excretory disease;
 KW urinary incontinence; renal disorder; neural; sensory disease;
 KW Alzheimer's disease; meningitis; respiratory disease; emphysema;
 KW occupational lung disease; endocrine disease; diabetes;
 KW glomerulonephritis; digestive disease; portal hypertension;
 KW irritable bowel syndrome; epithelial disease; scleroderma;
 KW epidermolysis bullosa; cytostatic; antineoplastic; antiarrhythmic;
 KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
 KW antiparkinsonian; antibacterial; osteopathic; dermatological; antitumor;
 KW immunomodulator; antiarrhythmic; cardiac; norepinephrine; antidiabetic;
 KW nephrotropic; uterine; neuroprotective; antiparkinsonian; tranquilizer;
 KW antidiabetic; anabolic; hypertensive; vulnary.
 OS Homo sapiens.
 PN US2004010132-A1.
 XX 15-JAN-2004.
 XX 30-OCT-2001; 2001US-00984429.
 XX 09-OCT-1997; 97US-0061463P.
 XX 09-OCT-1997; 97US-0061527P.
 XX 09-OCT-1997; 97US-0061529P.
 XX 09-OCT-1997; 97US-0061532P.
 XX 09-OCT-1997; 97US-0061536P.
 XX 09-OCT-1997; 97US-0071498P.
 XX 08-APR-1998; 98WO-US021142.
 XX 01-NOV-2000; 2000US-0244591P.
 XX (ROSE/) ROSEN C A.
 XX (BREW/) BREWER L A.
 XX (DUAN/) DUAN R D.
 XX (RUBEN/) RUBEN S M.
 XX (FLORENCE/) FLORENCE K A.
 XX (GREENE/) GREENE J M.
 XX (YOUNG/) YOUNG P E.
 XX (FERRIE/) FERRIE A M.
 XX (YUGG/) YU G.
 XX (FLORENCE/) FLORENCE C.
 XX (EBNER/) EBNER R.
 XX (OLSEN/) OLSEN H.
 PI Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM,
 PI Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;
 XX WPI; 2004-090518/09.
 DR N-PSDB; ADJ12161.
 XX New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating, preventing or ameliorating diseases or disorders e.g. cancer,
 PT anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's
 PT disease.
 XX Claim 11; SEQ ID NO 71; 286pp; English.
 XX This invention relates to novel polynucleotides encoding human secreted
 CC proteins. Specifically, it refers to the vectors, host cells, recombinant
 CC and synthetic methods for producing human polynucleotides, polypeptides
 CC and antibodies. Furthermore, it relates to screening methods to identify
 CC agonists and antagonists that can be used to inhibit or enhance the
 CC production and function of the secreted proteins. The present invention

CC describes these compositions as useful for diagnosing, treating or
 CC preventing disorders such as cancer, haematopoietic diseases including
 CC anaemia and multiple myeloma, reproductive system disorders including
 CC prostatitis and inguinal hernia, musculoskeletal diseases including
 CC systemic lupus erythematosus and gout, cardiovascular disease including
 CC arrhythmia and hypernatraemia, mixed fetal diseases including fetal
 CC alcohol syndrome and Down's syndrome, excretory diseases including
 CC urinary incontinence and renal disorders, neural or sensory disease
 CC including Alzheimer's disease and meningitis, respiratory diseases
 CC including emphysema and occupational lung disease, endocrine diseases
 CC including diabetes and glomerulonephritis, digestive diseases including
 CC portal hypertension and irritable bowel syndrome and connective tissue or
 CC epithelial diseases including scleroderma and epidermolysis bullosa. As
 CC such, there are various activities such as cytostatic, antineoplastic,
 CC antiarrhythmic, antiasthmatic, anti-HIV, immunosuppressive,
 CC antiinflammatory, antiparkinsonian, antibacterial, osteopathic,
 CC dermatological, antitumor, immunomodulator, antiarrhythmic, cardiac,
 CC norepinephrine, antidiabetic, anabolic, hypertensive and
 CC antiparkinsonian, tranquilizer, a human secreted protein of the invention.
 CC NOTE: This polypeptide is a human secreted protein of the invention.
 CC NOTE: This sequence does not appear in the printed specification but has
 CC been obtained in electronic format from the US patent office at the
 CC following web site www.seqdata.uspto.gov/sequence.html; Document ID:
 CC 20040010132.

XX Sequence 125 AA;

Query Match 18.0%; Score 568; DB 8; Length 125;

Best Local Similarity 100.0%; Pred. No. 1.5e-48; Indels 0; Gaps 0;
 Matches 107; Conservative 0; Mismatches 0;

QY 185 MRWILILVIALWFIELDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVITSL 244
 DB 1 MRWILILVIALWFIELDVWSTCSQPICAKWTREAEKSKLSSEGHMDLPDVITSL 60

QY 245 PGSGAEILKQLFFNSDFLIRVPTAYIDIPETELEDISFVDACEWK 291

DB 61 PGSGAEILKQLFFNSDFLIRVPTAYIDIPETELEDISFVDACEWK 107

RESULT 13

AA114415

ID AAY14415 standard; protein; 126 AA.

XX AAY14415;

DT 17-AUG-1999 (first entry)

DE Human secreted protein encoded by gene 5 clone HSABG21.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; chylod; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.

XX WO9919339-A1.

XX 22-APR-1999.

XX 08-OCT-1998; 98WO-US021142.

XX 09-OCT-1997; 97US-0061463P.

PR 09-OCT-1997; 97US-0061527P.

PR 09-OCT-1997; 97US-0061529P.

PR 09-OCT-1997; 97US-0061532P.

PR 09-OCT-1997; 97US-0061536P.

PR 09-OCT-1997; 97US-0071498P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Brewer LA, Olsen HS, Duan R, Ebner R, Rosen CA, Ruben SM;

PI Florence KA, Young PE, Greene JM, Yu G, Ferrie AM, Florence C;

XX DR N-PSDB; AAX79015.

XX PT New isolated human genes and the secreted polypeptides they encode.

XX PS Claim 11; Page 186; 226pp; English.

CC This sequence represents a secreted human protein encoded by the gene

CC clone detailed in the descriptor line. The gene can be used to generate

CC fusion proteins by linking to the gene to a human immunoglobulin Fc

CC portion (e.g. AAX79002) for increasing the stability of the fused protein

CC as compared to the human protein only. The invention relates to 53 novel

CC genes and their fragments (nucleic acid sequences: AAX79011-X79064; amino

CC acid sequences AAX14411-Y14464) which are useful for preventing, treating

CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,

CC pathological conditions can be diagnosed by determining the amount of the

CC new polypeptides in a sample or by determining the presence of mutations

CC in the new polynucleotides. Specific uses are described for each of the

CC 53 polynucleotides, based on which tissues they are most highly expressed

CC in (see AAX79011 for described uses)

XX SQ Sequence 126 AA;

Query Match 18.0%; Score 568; DB 2; Length 126;

Best Local Similarity 100.0%; Pred. No. 1.5e-48;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 MRWILLVIALWFIELDVNSTCSQPTCAKWTREAGSKSLSSGHHMDLPDVVITSL 244

DB 1 MRWILLVIALWFIELDVNSTCSQPTCAKWTREAGSKSLSSGHHMDLPDVVITSL 60

QY 245 PGSGAEILKQLFFNSDFLIRVPTAYIDIPETELEDVSFVDACEWK 291

DB 61 PGSGAEILKQLFFNSDFLIRVPTAYIDIPETELEDVSFVDACEWK 107

RESULT 14

ADN95522

ID ADN95522 standard; protein; 410 AA.

XX AC ADN95522;

XX DT 01-JUL-2004 (first entry)

XX DE Human BEC/LEC-related protein sequence SeqID445.

XX KW growth; differentiation; blood endothelial cell; BEC;

KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;

KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;

KW vasotrophic; antiinflammatory; gene therapy; endothelial cell disorder;

KW inflammatory disease; cancer metastasis; lymphatic system; human.

XX OS Homo sapiens.

XX PN WO2003080640-A1.

XX PD 02-OCT-2003.

XX PF 07-MAR-2003; 2003WO-US006900.

XX PR 07-MAR-2002; 2002US-0363019P.

XX PA (LUDW-) LUDWIG INST CANCER RES.

PA (LICN) LICENTIA LTD.

XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;

XX

DR WPI; 2003-876899/81.

XX N-PSDB; ADN95523.

PS Example 1; SEQ ID NO 445; 176pp; English.

XX This invention relates to a method of differentially modulating the

CC growth or differentiation of blood endothelial cells (BEC) or lymphatic

CC endothelial cells (LEC) comprises contacting endothelial cells with a

CC composition comprising an agent that differentially modulates blood or

CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises

CC identifying a human subject with lymphoedema and with a mutation in at

CC least one allele of a gene encoding a LEC protein, where the proviso that

CC the LEC protein is not VEGFR-3; and administering to the subject a

CC composition comprising a lymphatic growth agent selected from VEGF-C or

CC VEGF-D polypeptides and polynucleotides. The invention may be useful for

CC the development of compounds with an antiangiogenic, cytostatic,

CC vasotrophic or antiinflammatory activity or for gene therapy. The method

CC is useful in modulating the growth or differentiation of blood

CC endothelial cells or lymphatic endothelial cells, in treating hereditary

CC lymphoedema, in screening for an endothelial cell disorder or

CC predisposition to the disorder or in monitoring the efficacy or toxicity

CC of a drug on endothelial cells. The agent is useful in manufacturing a

CC medicament for the differential modulation of blood vessel endothelial

CC cell or lymphatic vessel endothelial cell growth or differentiation. The

CC lymphatic growth agent may also be used in manufacturing a medicament for

CC the treatment of hereditary lymphoedema resulting from a mutation in a

CC LEC gene or of other diseases involving the lymphatic vessels, such as

CC various inflammatory diseases and cancer metastasis via the lymphatic

CC system. The present sequence is that of a human LEC/BEC differentially

CC expressed protein which is related to the method of the invention. Note:

CC This sequence does not appear in the specification but was obtained by

CC the indexer using the source data given in table 14 of the specification.

XX SQ Sequence 410 AA;

Query Match 4.7%; Score 147; DB 7; Length 410;

Best Local Similarity 18.2%; Pred. No. 2.6e-05;

Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY 187 WILVIALWFIEL-----LDVWSTC-----SQPTCAKWTREAGSKSLSS 229

DB 5 WKAVLLALLASIAIOVTAIRTFKSFHTCPGLAELAEALRC-----ESPTFAYNLSR 59

QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLIRVPTAYIDIPETELEDVSFVDACE 289

DB 60 KTH-----ILLATRSGSSFFVQLFNQHLDFVLFEP----- 93

QY 290 WKVSDIRSGHFRLLRGWLSQVDTKLHLQNIHLHPEPNRGKLAQYFAMNKDKRKFRRRE 349

DB 94 -----YHVQNTLI PRFTOG-----KSPADRRV 115

QY 350 SUPERQSMKGAFDRDAEYI-----RALRHLVYPSARPV-----LS 387

DB 116 MLAGSRDLRLSYDCDLYPLENIKPPVPVNTDRIFRGASRVLCRSREVPDGPADLV 175

QY 388 LSSGSLTKLHFFQFVLGH-----SMREALYIVRD 416

DB 176 LEEGDCVRKCGLLNLTVAABACRSHVAIKTVRVPDVNDLRALVEDPRLNKVITQLVRD 235

QY 417 PRAWIYMLYNSKPSLYSLKQVPEHLAKLFKEGGKGCNLSNGYAFEYERKELSKSK 476

DB 236 PGILASRSETFRDITY-----RLWRLLWYGTGR-----KPNLDVTQL-TTVCEDF 279

QY 477 SNAVS--LLSHLWLANTAALRINTDLPTSQVLFKEDIHVPFQKTTTRIFAFGLIPL- 533

DB 280 SNSVSTGLMRPPLKLG-----KYMVLRYEDLARPNMKKTEIYGLGIPLD 325

QY 534 -----SPASLNQILFATSTNLPFLYPEGEISPTNTNVKQNI-PROEIKLIENI 581

DB 326 SHVARWIONTRGDPTLGHKHYGTVRN-----SAATAEKWRFRLSYDIVAFQA 375

QY 582 CWTLMRLCY 591

Search completed: June 23, 2005, 08:43:25
Job time : 66.4109 secs

Sequence 411 AA:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 22.4643 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-9

Perfect score: 3147

Sequence: 1 MNRNGAMVDVDAHYKMFV.....LIENICWLMRLGYPKEMD 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	568	18.0	126	4	US-09-288-143-71
2	147	4.7	411	3	US-09-015-188-2
3	134	4.3	483	3	US-09-263-023-2
4	134	4.3	483	4	US-09-471-867-2
5	133	4.2	484	3	US-09-263-023-4
6	133	4.2	484	4	US-09-471-867-4
7	133	4.2	531	4	US-09-949-016-6471
8	133	4.2	608	4	US-09-949-016-9449
9	123.5	3.9	395	4	US-09-949-016-7011
10	120.5	3.8	479	2	US-08-895-514-2
11	119.5	3.8	390	4	US-09-949-016-6813
12	119.5	3.8	431	4	US-09-949-016-8893
13	115.5	3.7	386	4	US-09-786-240-11
14	108.5	3.4	386	3	US-09-045-284A-2
15	108.5	3.4	386	3	US-09-190-911-1
16	108	3.4	826	4	US-09-252-991A-22143
17	103.5	3.3	324	4	US-09-270-767-43224
18	103	3.3	582	3	US-09-091-725-17
19	101.5	3.2	594	3	US-08-987-151-2
20	101	3.2	650	4	US-09-248-796A-16387
21	99.5	3.2	1060	4	US-09-248-796A-15476
22	99.5	3.2	601	4	US-09-107-433-3186
23	99	3.1	376	3	US-09-150-133-7
24	99	3.1	376	3	US-09-150-141-7
25	99	3.1	376	3	US-09-374-493-7
26	99	3.1	376	3	US-09-374-824-7
27	99	3.1	376	3	US-09-374-492-7

28	99	3.1	376	4	US-09-785-343-7	Sequence 7, Appli
29	99	3.1	376	4	US-10-411-976-7	Sequence 7, Appli
30	98	3.1	591	4	US-09-583-110-3756	Sequence 3756, Ap
31	97.5	3.1	1172	4	US-09-328-352-6071	Sequence 6071, Ap
32	97	3.1	420	4	US-09-949-016-6705	Sequence 6705, Ap
33	97	3.1	776	4	US-09-949-016-7231	Sequence 7231, Ap
34	96.5	3.1	843	4	US-09-248-796A-17210	Sequence 17210, A
35	96	3.1	458	2	US-08-655-878-2	Sequence 2, Appli
36	94	3.0	232	4	US-09-248-796A-16863	Sequence 16863, A
37	94	3.0	421	4	US-09-949-016-8165	Sequence 8165, Ap
38	94	3.0	756	4	US-09-919-497-66	Sequence 66, Appl
39	93	3.0	888	4	US-09-538-092-787	Sequence 787, App
40	91.5	2.9	721	4	US-08-851-435-2	Sequence 2, Appli
41	91.5	2.9	746	4	US-08-851-435-6	Sequence 6, Appli
42	91.5	2.9	915	4	US-09-583-110-3100	Sequence 3100, Ap
43	91	2.9	365	4	US-09-818-512-4	Sequence 4, Appli
44	91	2.9	370	3	US-09-150-133-1	Sequence 1, Appli
45	91	2.9	370	3	US-09-150-141-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-288-143-71
; Sequence 71, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-09-288-143-71

Query Match 18.0%; Score 568; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 2e-53;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	185	MRWLLIIVIALWFLFDVMTCSQPCAKWTRTEAGSKLSSEGHMDLPDVTSL	244
Db	1	MRWLLIIVIALWFLFDVMTCSQPCAKWTRTEAGSKLSSEGHMDLPDVTSL	60
Qy	245	PGSGAEILKQLFFNSSDFLIRVPTAYIDIPETELEIDSFVDACEWK	291
Db	61	PGSGAEILKQLFFNSSDFLIRVPTAYIDIPETELEIDSFVDACEWK	107

RESULT 2


```
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 4.2%; Score 133; DB 3; Length 484;
Best Local Similarity 26.2%; Pred. No. 4.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 279 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 337
Qy 460 GYAFEYELRKELSKSNVLSLHLWNTAAALRINTDLPSTSYOLVKFEDIVHFPQ 519
Db 338 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHVYLVRYEDLVGDVP 387
Qy 520 KTERIFAFGLIPLSA-----SLNQILFATSTNLFYPYGEISPTN-----TNVWKQNL 571
Db 388 KTLRRVDFVGLLVSPENEFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTLMRLGYPK 593
Db 443 FQIKQVEEFYQPMVILGYER 464

RESULT 7
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-6471

Query Match 4.2%; Score 133; DB 4; Length 531;
Best Local Similarity 26.2%; Pred. No. 5e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 326 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 384
```

```
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAM41.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-471-867-4

Query Match 4.2%; Score 133; DB 4; Length 484;
Best Local Similarity 26.2%; Pred. No. 4.3e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 279 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 337
Qy 460 GYAFEYELRKELSKSNVLSLHLWNTAAALRINTDLPSTSYOLVKFEDIVHFPQ 519
Db 338 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHVYLVRYEDLVGDVP 387
Qy 520 KTERIFAFGLIPLSA-----SLNQILFATSTNLFYPYGEISPTN-----TNVWKQNL 571
Db 388 KTLRRVDFVGLLVSPENEFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTLMRLGYPK 593
Db 443 FQIKQVEEFYQPMVILGYER 464

RESULT 7
US-09-949-016-6471
; Sequence 6471, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6471
; LENGTH: 531
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Human
US-09-949-016-6471

Query Match 4.2%; Score 133; DB 4; Length 531;
Best Local Similarity 26.2%; Pred. No. 5e-05;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAWYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
Db 326 LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRAHRMPPFLEAAGHKLGAKEGV 384
```

[illegible]

RESULT 8
US-09-949-016-9449
; Sequence 9449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9449

Query Match	4.2%;	Score 133;	DB 4;	Length 608;
Best Local Similarity	26.2%;	Pred. No. 6.3e-05;		
Matches 53;	Conservative 36;	Mismatches 81;	Indels 32;	Gaps 9;
Qy	408	MRALYIVRDPRAWTYSMLYNKSPLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS	459	
Db	403	LKVIHLVRDPA-VASSRISSRHCLIRBSLQVRSRDPRAHRMPLPBAAGHLGAKGKV	461	
Qy	460	GYAFEPYELRKELSKSNVALLSHLWLANTAALINTDLPTSQYLVKFEDIVIHFPQ	519	
Db	462	GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLGCHYLVVRVEDLVGDPV	511	
Qy	520	KTTERIFAFGLGIPLSPA-----SLNQILFATSTNLFLPYEGEISPTN----TNVWKQNL	571	
Db	512	KTLRVRVDFGLLVSPNMEQFALNTSGSGSSKPFV-----VSARNATQAAANWRTALT	566	
Qy	572	RDEIKLIENICWTLMDLRGYPK	593	
Db	567	FOOI KOVEEFCYOPMAVLGVGR	588	

RESULT 9
US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

```

; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: PaatSEQ for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7011

Query Match          3.9%; Score 123.5; DB 4; Length 395;
Best Local Similarity 23.3%; Pred. No. 0.00032;
Matches 90; Conservative 53; Mismatches 140; Indels 103; Gaps 20;

QY 239 VVITSLPGSGAIIKOLFENSSDFLIYIRVPTAYIDIPETELEIDSFVDACEWKV-SDIRS 297
Db 43 VIVLSWSRSGSFVGLFQNHQHPDVFLMEPA-----WHVWTTLSQ 82

QY 298 GHFRLLRGLWQLSVQDTKL-HLQNIHLHNPNGKLAQYFAMNKDKQKFKRRESLPEQRS 356
Db 83 GSAATLHMAVRDLVRSVFLCDMDVFDALPWRNLSDLF-----QWAVSRLCSPPACS 136

QY 357 QMKGAFDRDAEYIRALRHLVYYS--ARPVLSGSGSWTLKLHFTQ-EVL-----G 405
Db 137 ----APPRGAISSEAVCKPLCARQSFITLAREACRSYSHVVLKEVFFNLQVLPILLSDA 192

QY 406 ASMRALYIVDRPAMIYMLNKPSSLYSLKNVPEHLAKLFIEGG--KGKCNLNSGYAF 463
Db 193 LNLRIVHLVRDPRAVLRSR-----EQTAKALARONGIVLG---TNGTWTV 233

QY 464 EYEP-LR--KELSKSNAVALSHLWLTAAALRINTDLLPTS YQLVKPEDIWHFPQK 520
Db 234 EADPGLRVVRECR-----SHVRIAE-AATLK-PPPLRGYRLVRFEEDLAREPLA 282

QY 521 TTERIPAFGLIPSP---ASLNQI-----LFATSTNLPFLPYEGEISPTN-TNV 565
Db 283 EIRALYAFGLSLTQLEAWIHNTHGSGPGARREAPKTSR-----NALNVSOA 332

QY 566 WQNQLPRDEIKLIENICWTLMDRIGY 591
Db 333 WRHALPPAKIRRVQBELCAGALQLLGY 358

RESULT 10
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-514-2

Query Match      3.8%; Score 120.5; DB 2; Length 479;
Best Local Similarity 16.7%; Pred. No. 0.00096;
Matches 71; Conservative 64; Mismatches 152; Indels 137; Gaps 12;

QY      217 RTEAGSKKSLSSGHHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP----- 268
Db      118 RKEEPPRPVAGPRRH-----VLLMATRTGSSVGFGEFFNQOQNIPLYFEPLMHIERTV 172
QY      269 -----TAYIDIPETELEIDSFVDACEWKVSDIRSGH---FLLRGWLQSLVQ 312
Db      173 SFEPGGANAAGSALVYRDVLKQLFLCDLYV--LEHFTITPLPEDHLTQFMFRGSSRLCE 230
QY      313 D-----TKLHLQNIHLHEPNRGKLAQYFAMNKDKKRRKRESLPQSRQOMKGAFDRDA 366
Db      231 DPVCTPPVKVFEKYCHCKNRCGGLNVTLAAEACRKK-----EHWALKAVRIQOL 280
QY      367 EYRALRRHLVYPSAPVLSLSSGSWTLKLHFFQEVGLGSMRALYIVRDPRAWIYSMLY 426
Db      281 EFLQPL-----AEDPRLDL-----RVQLVRDPRAVLASRMV 312
QY      427 NSKPSLYSLKNVPEHLAKLFKEGGKGCNLSNGVAPEYEPRLRKELSKSKSNVLSLHL 486
Db      313 -----APAGKYTKWKWLDDGODGUREEVQ 339
QY      487 WLANTAALRINTDL-----LPTSQYLVKFEIVHFPQKTTERRIFAFGLGIPLSASLN 539
Db      340 RLRCNCESIRLSAELRQPAWLGRYMLVRYEDVARGPLQKAREMYFFAGIPLTPQVED 399
QY      540 QI-----LFTSTNLFYLPYEGETSPNTNWKQNLPRDEIKLIENICWTLM 587
Db      400 WIQKNTQAHDGSGIYSTQKN-----SSQFEKWFRESMFPKLAQVVQAPCGPAMR 449
QY      588 RLGY 591
Db      450 LFGY 453

RESULT 11
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

Query Match      3.8%; Score 119.5; DB 4; Length 431;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY      223 SKKSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
Db      69 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 128
QY      271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326

; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6813

Query Match      3.8%; Score 119.5; DB 4; Length 390;
Best Local Similarity 21.6%; Pred. No. 0.00086;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY      223 SKKSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
Db      28 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 87
QY      271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326
Db      88 TLHMAVRDLMRISIFL--CMDVDFDAYMPQSRNLSAFENWATSRA-----LCSPP 134
QY      327 -----NRGKLAQYFAMNKDKKRRKRESLPQSRQOMKGAFDRDAEYIRALRRHLVYPS 381
Db      135 ACSAFPRTGTSK-----QDVCKTLCRTQPFSLAREACRSYSHVVVLKEVRFNFNQLVPL 189
QY      382 ARPVLSSSGSWTLKLHFFQEVGLGSMRALYIVRDPRAWIYSMLYSKPSLYSLKNVPEH 441
Db      190 SDPALNL-----RIVHLVRDPRAVLRSR--EAAAGPILARDNGI--- 225
QY      442 LAKLFKEGGKGCNLSNGVAPEYEP---LRKELSKSKSNVLSLHLWLAATAALRN 498
Db      226 -----VLGTNGK-----VEADPHLRILREVCR-----SHVRIAE--AATLK-P 261
QY      499 TDLPTSQYLVKFEIVHFPQKTTERRIFAFGLGIPLSASLNQILFATSTNLFYLPYEG-E 557
Db      262 PFLRGYRLVRPFDLAREPLAIBIRALYAFDTGLTLPQLEAWIHNITHGSGIGKPIEAFH 321
QY      558 ISPTN-----TNWKQNLPRDEIKLIENICWTLM 591
Db      322 TSSRNARNVSNQAWRHAPFTTKILRVQEVCAQALQLGY 359

RESULT 12
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

Query Match      3.8%; Score 119.5; DB 4; Length 431;
Best Local Similarity 21.6%; Pred. No. 0.001;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

QY      223 SKKSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVP-----TA 270
Db      69 SRPGSPAGGEDRVHVLVLSWSRSGSFLGQFSDHPDVFYLMPEPAHWVTTLSQSSAA 128
QY      271 YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP- 326

```

```

129  TLHMAVRDLMRRIFU--CDMDVFDAYMQSNLSAFFNWATSR-----ICSP 175
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
327  -----NRGLAQYFAMNKDKKFKFRRESLPEQSQMKGAFDRDAEYIRALRRHLVYYP 381
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
176  ACSAPFRGTISK-----QDVCKTLCTQPPSLAREACRSYSHVVLKEVRFNLOVLP 230
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
382  ARPVLSSGSGWTCLKHFQFVLGASMRALYIVRDPRAWIYSMLYNSKPSLSLKNVP 441
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
231  SDPALNL-----RIVHLVRDPRAVLRSR--EAAGPILARDNGI--- 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
442  LAKLFKEGGKGNLNGYAFVEP---LRKELSKSNVSLSHLMLANTAAALRN 498
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
267  -----VLGTNGK-----VEADPHURLIREVCR-----SHVRIAE--AATLK-P 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
499  TDLPTSVLKVFEDIVHPQKKTTERIFAFGLIPLSPASLNQILFATSTNLFVPEG-E 557
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
303  PPLRGRLVLPFEDLAREPLAEIRALYVAFGLTLTQLEAWIHNTHSGICKPIEAPH 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
558  IGPTN-----TNWKNQLPRDBIKLIENICWTLMDRLG 591
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
363  TSSRNARNVSQWRHALPFTKILRIVQECAGALQLG 400
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. Tom
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: AZIMZAI, Yalda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

```

	Query Match	3.7%;	Score 115.5;	DB 4;	Length 386;
	Best Local Similarity	20.4%;	Pred. No. 0.0023;		
	Matches	83;	Conservative	56;	Mismatches 151; Indels 117; Gaps 17
Qy	223 SKKSJSSSEGHMDLPDVPVITSLPGSGAAILKQLFFNSDFLYIRVP-----TAY 271	:	:	:	:
Dd	32 SSLSKMAQPERM---HVLVLTSWSRSGSFGQLFGQHPDVFLYLMPEAWHVMTTFKQSTAW 88	:	:	:	:
Qy	272 IDIPETELEIDSFVDA---CSWKYSDI-----RSQHFRLLRGWLQSLVDOTKLHLQN 320	:	:	:	:
Dd	89 M----LHWAVRDILRAVFLCDMSVFDAYMEGPPRSSLFQ-----WENSRALCSPACDI 140	:	:	:	:
Qy	321 IHLHPNRGLKAQYFAMNKKRKFKRESPLPQRSOMKGAFDRDAEYTRALRRHLVYY- 379	:	:	:	:
Dd	141 IPQDESSPGLTAGSCAVNPSPLKLEKACS-----YSHVVILKEVRFFEN 183	:	:	:	:
Qy	380 -PSARPVLSSLSSGSWTLKHPFOEVLGASMRALYIVRPRAWIYSMLYNKSPLSYSLKNV 438	:	:	:	:

```

Db      184  LQSLYPELLKDPSS---LNLHI-----VHLVRDPA-----VFRSR----- 214
Qy      439  PEHLAKLFKIEGGKGNKLSNGYAF--EYEPLRKELSKSKSNVLSLLSHLWLANTAALR 496
Db      215  -----ERTKGLMIDRSIRVMGQHEQKLKKEQDPYVMQVICQSQLEIYKTIQSL- 263
Qy      497  INTDLLPTSVQLVKFEDIVHFPOKTTTERIPAFGLIGPLSPASLNQILFAT-----STNLFY 551
Db      264  --PKALQERYLLVRYEDLARAPVAQTSRMVEFVGLFPLPHLOTVWNHNTGKMGGDHAFH 321
Qy      552  LPVEGEISPTN-----TNVWKONLPDREIKLIENICWTLMORLGY 591
Db      322  -----TNARDALNVQAWRWSLDPYEKVSRLQKACGDAMNLLGY 359

RESULT 14
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match      3.4%; Score 108.5; DB 3; Length 386;
Best Local Similarity 18.9%; Pred. No. 0.013;
Matches 78; Conservative 55; Mismatches 151; Indels 129; Gaps 15;

Qy      223  SKKSLSEGHHMDLPDVITSLPGSGAETLKQLFFNSSDFLYIRVPTAYIDIPETELEID 282
Db      32  SSLSMKAQPERM---HVLVLSSWRSGSSFGVQFGQHPDVFYLMPEPAWHV----- 78
Qy      283  SFVDACEWKVSDIRSGHFRLLRGWLSQVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKK 342
Db      79  -----WMT-----FKQSTAWM-----LHMAVRDLIRAVFLCDMSVF 109
Qy      343  RKFKRRRESLPEQASQWKGAFFDRDAEYIRAL-----RRHLVYTPSARPVL 386
Db      110  DAY--MEPPRRGSSL-----FOWENSRLACAPACDIIPODEIIPRAHCRLLCSQPPE 162
Qy      387  SLSSGSWTLKLHFFQVILGASBREALY-----IVRDPRAWIYSMLYNSKPSL 432
Db      163  VBEKACRSYHVVLKVEFRFNLOSPLPLKDPSELNLIHVLVRDPRA-----VFRSR--- 214
Qy      433  YSLKNVPEHLAKLFKIEGGKGNKLSNGYAF--EYEPLRKELSKSKSNVLSLLSHLWLAN 490
Db      215  -----ERTKGLMIDRSIRVMGQHEQKLKKEQDPYVMQVICQSQLEIYK 258
Qy      491  TAAALINTDLLPTSQVLVKFEDIVHFPOKTTTERIPAFGLIGPLSPASLNQILPAT----- 545
Db      259  TIQSL---PKALQERYLLVRYEDLARAPVAQTSRMVEFVGLFPLPHLOTVWNHNTGKMG 315
Qy      546  STNLFYLPVEGEISPTN-----TNVWKONLPDREIKLIENICWTLMORLGY 591
Db      316  GDHAFH-----TNARDALNVQAWRWSLDPYEKVSRLQKACGDAMNLLGY 359

RESULT 15
US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:

```

Search completed: June 23, 2005, 08:52:13
Job time : 24.4643 secs

THIS PAGE BLANK (USPTO)

Query Match	100.0%;	Score 3147;	DB 16;	Length 596;
Best Local Similarity	100.0%;	Prod. No. 2.Se-295;		
Matches 596;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	1	MNRVNGAMDVDWAHYKQFVFDHNGNSPMASIQEAEQAQAFKKKCTQFVNVTFWQESRIT	60	
Db	1	MNRVNGAMDVDWAHYKQFVFDHNGNSPMASIQEAEQAQAFKKKCTQFVNVTFWQESRIT	60	
Qy	61	RIAPVYFGPYINSSCCRFIDSSNPGLCISLVNNTTEHVSVITVDYHNLKTRFNYLGFPGF	120	

Db 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Qy 121 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 180
Db 121 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 180
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSGHMDLPDVV 240
Db 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSGHMDLPDVV 240
Qy 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
Db 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
Qy 301 RLLRGWLQSLVQDTKLHQLNHLHPEHNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Db 301 RLLRGWLQSLVQDTKLHQLNHLHPEHNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Qy 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Db 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Qy 421 IYSMLYNSKPSLSLXKNVPEHLAKLFKEGGKGCNLSNGYAFEYEPRLKELSKSNAV 480
Db 421 IYSMLYNSKPSLSLXKNVPEHLAKLFKEGGKGCNLSNGYAFEYEPRLKELSKSNAV 480
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPFEDIVHFPQKTTTERIFAFGLIGPLSPASLNQ 540
Db 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPFEDIVHFPQKTTTERIFAFGLIGPLSPASLNQ 540
Qy 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 596
Db 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 596

RESULT 2
; Sequence 15, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138Div
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-15

Query Match 100.0%; Score 3147; DB 16; Length 1222;
Best Local Similarity 100.0%; Pred. No. 7, 6e-295;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDVDAHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVTFQMESTIT 60
Db 627 MNRNGAMVDVDAHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVTFQMESTIT 686
Qy 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Db 687 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 746
Qy 121 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 180

Db 747 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 806
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSGHMDLPDVV 240
Db 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSGHMDLPDVV 866
Qy 241 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
Db 867 ITSLPGSGABILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 926
Qy 301 RLLRGWLQSLVQDTKLHQLNHLHPEHNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 360
Db 927 RLLRGWLQSLVQDTKLHQLNHLHPEHNRGKLAQYFAMNKDKKFKRRESLPRQSQMKG 986
Qy 361 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 420
Db 987 AFDRDAEYIRALRRHLVYPSARPVLSLSSGSWTLKLHFFQEVLGASMRALYIVRDPRAW 1046
Qy 421 IYSMLYNSKPSLSLXKNVPEHLAKLFKEGGKGCNLSNGYAFEYEPRLKELSKSNAV 480
Db 1047 IYSMLYNSKPSLSLXKNVPEHLAKLFKEGGKGCNLSNGYAFEYEPRLKELSKSNAV 1106
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPFEDIVHFPQKTTTERIFAFGLIGPLSPASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKPFEDIVHFPQKTTTERIFAFGLIGPLSPASLNQ 1166
Qy 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 596
Db 1167 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 1222

RESULT 3
US-10-479-472A-2
; Sequence 2, Application US/10479472A
; Publication No. US20050118581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LOE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-479-472A-2

Query Match 99.8%; Score 3142; DB 17; Length 1212;
Best Local Similarity 99.8%; Pred. No. 2, 3e-294;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNRNGAMVDVDAHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVTFQMESTIT 60
Db 617 MNRNGAMVDVDAHYKMFWDHGHGNSPMASIQEAEQAAEFKKRWTQFVNVTFQMESTIT 676
Qy 61 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 120
Db 677 RIAYVFGPYINSSCRFIDSSNPGLOISLVNNTTEHWSIVTDYHNLKTRFNVLGFGF 736
Qy 121 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 180
Db 737 ASVADQOITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLILCISLVILTTFQWRPFLS 796
Qy 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSSSGHMDLPDVV 240

Db 797 FRKLMRWILILVIALWFELDDVWSTCSQPICAKWTREAGSKSLSSSEGHMDLPDV 856
Qy 241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 300
Db 857 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 916
Qy 301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKRKFKRRRESLPEQRSQMG 360
Db 917 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKRKFKRRRESLPEQRSQMG 976
Qy 361 AFDDAEYIRALRHLYVYPSRVLSSSGSWTLKHLFFQEVLGASNRALYIYVRDPRAW 420
Db 977 AFDDAEYIRALRHLYVYPSRVLSSSGSWTLKHLFFQEVLGASNRALYIYVRDPRAW 1036
Qy 421 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGNLNSGVAFYEPYLRKELSKSNAV 480
Db 1037 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGNLNSGVAFYEPYLRKELSKSNAV 1096
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLGIPLSPASLNQ 540
Db 1097 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLGIPLSPASLNQ 1156
Qy 541 ILFATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENICWTLMDRLGYPKPM 596
Db 1157 ILFATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENICWTLMDRLGYPKPM 1212

RESULT 4

US-10-475-446-4
; Sequence 4, Application US/10475446
; Publication No. US20040198651A1
; GENERAL INFORMATION:
; APPLICANT: KLAMMER, Aaron A.; HAFALIA, April J.A.
; APPLICANT: DUGAN, Brendan M.; WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.; TRIBOULEY, Catherine M.
; APPLICANT: ARVIZU, Chandra S.; HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Dannel B.; KALLICK, Deborah A.
; APPLICANT: YUE, Henry; AU-YOUNG, Janice K.
; APPLICANT: RAKUMAR, Jayalaxmi; LI, Joana X.
; APPLICANT: THANGAVELU, Kavitha; GIETZEN, Kimberly J.
; APPLICANT: DING, Li; BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.
; APPLICANT: MASON, Patricia M.; LAL, Preeti G.
; APPLICANT: GRAUL, Richard C.; REDDY, Roopa M.
; APPLICANT: BECHA, Shanya D.; KAREHT, Stephanie K.
; APPLICANT: RICHARDSON, Thomas W.; TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; APPLICANT: XU, Yuming
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PF-0949 USN
; CURRENT APPLICATION NUMBER: US/10/475,446
; PRIOR FILING DATE: 2003-10-20
; CURRENT APPLICATION NUMBER: PCT/US02/12464
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,207
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/287,114
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/288,640
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/290,516
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/292,184
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/343,553
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/357,002
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 60/358,279
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: US 60/366,041
; PRIOR FILING DATE: 2002-03-19

; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 3441255CD1
US-10-475-446-4

Query Match 99.8%; Score 3142; DB 16; Length 1222;

Best Local Similarity 99.8%; Pred. No. 2.3e-294;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MNRVNGAMDMVDAHYKMFWDHGHNSPMASIOBAEQAAEPKRWTOFVNVTFFOMESIT 60
Db 627 MNRVNGAMDMVDAHYKMFWDHGHNSPMASIOBAEQAAEPKRWTOFVNVTFFOMESIT 686
Qy 61 RIAYVFYGPYINVSSCRFIDSSNPGLQISLVNNTNTEHVSIIVTDYHNLKTRFNVLGPGF 120
Db 687 RIAYVFYGPYINVSSCRFIDSSNPGLQISLVNNTNTEHVSIIVTDYHNLKTRFNVLGPGF 746
Qy 121 ASVADQSQITRFGLTQAIKVPVRHRIIPFGPKFNIAVGLILCISLVILTFQWRFYLS 180
Db 747 ASVADQSQITRFGLTQAIKVPVRHRIIPFGPKFNIAVGLILCISLVILTFQWRFYLS 806
Qy 181 FRKLMRWILILVIALWFELDDVWSTCSQPICAKWTREAGSKSLSSSEGHMDLPDV 240
Db 807 FRKLMRWILILVIALWFELDDVWSTCSQPICAKWTREAGSKSLSSSEGHMDLPDV 866
Qy 241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 300
Db 867 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDCACWKVSDIRSGHF 926
Qy 301 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKRKFKRRRESLPEQRSQMG 360
Db 927 RLLRGWLQSLVQDTKLHLQNLHLEPNRGKLAQYFANMKDKRKFKRRRESLPEQRSQMG 986
Qy 361 AFDDAEYIRALRHLYVYPSRVLSSSGSWTLKHLFFQEVLGASNRALYIYVRDPRAW 420
Db 987 AFDDAEYIRALRHLYVYPSRVLSSSGSWTLKHLFFQEVLGASNRALYIYVRDPRAW 1046
Qy 421 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGNLNSGVAFYEPYLRKELSKSNAV 480
Db 1047 IYSLMYSKPSLSYSLKNVPEHLAKLFKEGGKGNLNSGVAFYEPYLRKELSKSNAV 1106
Qy 481 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLGIPLSPASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHFPQKTERIFAFGLGIPLSPASLNQ 1166
Qy 541 ILFATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENICWTLMDRLGYPKPM 596
Db 1167 ILFATSTNLFYLPYEGEISPTNTNWKQNLPRDEIKLIENICWTLMDRLGYPKPM 1222

RESULT 5

US-10-697-828-17
; Sequence 17, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: mouse
; US-10-697-828-17

Query Match          90.1%; Score 2835; DB 16; Length 1207;
Best Local Similarity 88.6%; Pred. No. 1.3e-264;
Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDADHYKMFDFDHGNSPMASIOEAOAAEFKKRWTFQVNVVTFQMESTIT 60
DB 612 MNRNGAMDMVDADHYKMFDFDHGNSPMVANOIEAOAAEFKKRWTFQVNVVTFQMESTIT 671

QY 61 RIAYVFGPYNVSSCRFDIDSSNPGIQISLVNNTTEHVSVITDYHNLKTRFNYLGFGGF 120
DB 672 RIAYVFGPYNVSSCRFDIDSSSGLQISLVNSTEHSVSVDYQNLKSRFSLGFGGF 731

QY 121 ASVADOGITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLICISLVILTTQWRPFLS 180
DB 732 ASVANOQITRFGIGTQEIIVNVRHDKVNPFPFGFKFNIAVGFILCISLVILTTQWRPFLS 791

QY 181 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSSGHHMDLPDVV 240
DB 792 FRKLMRCVLIVIALWFIELLDVWSTCTQPICAKWTREAKANEKVMISGHHVDLPNVI 851

QY 241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
DB 852 ITSLPGSGAEILKOLFNSDFLYIRIPTAYMDIPETEFEIDSFVDACEWKVSDIRSGHF 911

QY 301 RLLRGWLSQVQDTKLHQLNIHLHFNPKLAQYFAMNKDKKFKRRESLPQRSOMKG 360
DB 912 HLLRGWLSQVQDTKLHQLNIHLHETSRKLAQYFTTNKDKKFKRRESLQQRSRIG 971

QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
DB 972 PFDRDAEYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGTSMRALYIVRDPRAW 1031

QY 421 IYSLMYSKPSLSYLNKVPHEHLAKFKIEGGKGCNLSNGYAFEPYELPKELSKSNAV 480
DB 1032 IYSLVYGSFESLSYLNKVPHEHLAKFKIEGSKCNSNGYAFEYESLKELEISQNAI 1091

QY 481 SLLSHLWLANTAALINTDLPSTYQLVKFEDIVHFPQKTTERRIFAFIGIPLSPASLNQ 540
DB 1092 SLLSHLWVANTAALINTDLPNTYHLVKFEDIVHFPQKTTERRIFAFIGIPLSPASLNQ 1151

QY 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 1152 MLFATSTNLFYLPYEGEISPSNTNIWKTNLPRDEIKLIENICWTLMDHLGYPKFMD 1207

RESULT 6
US-10-479-472A-4
; Sequence 4, Application US/10479472A
; Publication No. US2005018581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479, 472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 4
; LENGTH: 1207
; TYPE: PRT

```

```

; ORGANISM: Mus sp.
; US-10-479-472A-4

Query Match          90.1%; Score 2835; DB 17; Length 1207;
Best Local Similarity 88.6%; Pred. No. 1.3e-264;
Matches 528; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

QY 1 MNRNGAMDMVDADHYKMFDFDHGNSPMASIOEAOAAEFKKRWTFQVNVVTFQMESTIT 60
DB 612 MNRNGAMDMVDADHYKMFDFDHGNSPMVANOIEAOAAEFKKRWTFQVNVVTFQMESTIT 671

QY 61 RIAYVFGPYNVSSCRFDIDSSNPGIQISLVNNTTEHVSVITDYHNLKTRFNYLGFGGF 120
DB 672 RIAYVFGPYNVSSCRFDIDSSSGLQISLVNSTEHSVSVDYQNLKSRFSLGFGGF 731

QY 121 ASVADOGITRFGIGTQAIKVPVRHRIIPFGFKFNIAVGLICISLVILTTQWRPFLS 180
DB 732 ASVANOQITRFGIGTQEIIVNVRHDKVNPFPFGFKFNIAVGFILCISLVILTTQWRPFLS 791

QY 181 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSSGHHMDLPDVV 240
DB 792 FRKLMRCVLIVIALWFIELLDVWSTCTQPICAKWTREAKANEKVMISGHHVDLPNVI 851

QY 241 ITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300
DB 852 ITSLPGSGAEILKOLFNSDFLYIRIPTAYMDIPETEFEIDSFVDACEWKVSDIRSGHF 911

QY 301 RLLRGWLSQVQDTKLHQLNIHLHFNPKLAQYFAMNKDKKFKRRESLPQRSOMKG 360
DB 912 HLLRGWLSQVQDTKLHQLNIHLHETSRKLAQYFTTNKDKKFKRRESLQQRSRIG 971

QY 361 AFDRDAEYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGASMRALYIVRDPRAW 420
DB 972 PFDRDAEYIRALRRHLVYPSARPVLSSGSGWTLKLFHFFQEVLGTSMRALYIVRDPRAW 1031

QY 421 IYSLMYSKPSLSYLNKVPHEHLAKFKIEGGKGCNLSNGYAFEPYELPKELSKSNAV 480
DB 1032 IYSLVYGSFESLSYLNKVPHEHLAKFKIEGSKCNSNGYAFEYESLKELEISQNAI 1091

QY 481 SLLSHLWLANTAALINTDLPSTYQLVKFEDIVHFPQKTTERRIFAFIGIPLSPASLNQ 540
DB 1092 SLLSHLWVANTAALINTDLPNTYHLVKFEDIVHFPQKTTERRIFAFIGIPLSPASLNQ 1151

QY 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
DB 1152 MLFATSTNLFYLPYEGEISPSNTNIWKTNLPRDEIKLIENICWTLMDHLGYPKFMD 1207

RESULT 7
US-09-984-429-71
; Sequence 71, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,591
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536

```

;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/061,532
;; PRIOR FILING DATE: 1997-10-09
;; NUMBER OF SEQ ID NOS: 727
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 71
;; LENGTH: 125
;; TYPE: .PRT
;; ORGANISM: Homo sapiens
US-09-984-429-71

Query Match 18.0%; Score 568; DB 11; Length 125;
Best Local Similarity 100.0%; Pred. No. 1.6e-46;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 185 MRWLLIVIALWFLLDVWSTCSQPCAKWTRTEAGSKLSLSEGHMDLPDVVITSL 244
Db 1 MRWLLIVIALWFLLDVWSTCSQPCAKWTRTEAGSKLSLSEGHMDLPDVVITSL 60

Qy 245 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 291
Db 61 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 8
US-10-150-111-71
;; Sequence 71, Application US/10150111
;; Publication No. US20030078386A1
;; GENERAL INFORMATION:
;; APPLICANT: Rubin et al.
;; TITLE OF INVENTION: Secreted Protein HPEAD48
;; FILE REFERENCE: P2018P1D1
;; CURRENT APPLICATION NUMBER: US/10/150.111
;; CURRENT FILING DATE: 2002-05-20
;; PRIOR APPLICATION NUMBER: 09/288,143
;; PRIOR FILING DATE: 1999-04-08
;; PRIOR APPLICATION NUMBER: PCT/US98/21142
;; PRIOR FILING DATE: 1998-10-08
;; PRIOR APPLICATION NUMBER: 60/061,463
;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/061,529
;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/071,498
;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/061,527
;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/061,536
;; PRIOR FILING DATE: 1997-10-09
;; PRIOR APPLICATION NUMBER: 60/061,532
;; PRIOR FILING DATE: 1997-10-09
;; NUMBER OF SEQ ID NOS: 219
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 71
;; LENGTH: 126
;; TYPE: .PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (126)
;; OTHER INFORMATION: Xaa equals stop translation
US-10-150-111-71

Query Match 18.0%; Score 568; DB 14; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.6e-46;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 185 MRWLLIVIALWFLLDVWSTCSQPCAKWTRTEAGSKLSLSEGHMDLPDVVITSL 244
Db 1 MRWLLIVIALWFLLDVWSTCSQPCAKWTRTEAGSKLSLSEGHMDLPDVVITSL 60

Qy 245 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 291
Db 61 PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 9
US-10-021-660-128
;; Sequence 128, Application US/10021660
;; Publication No. US20030152926A1
;; GENERAL INFORMATION:
;; APPLICANT: Murray, Richard
;; APPLICANT: Glynn, Richard
;; APPLICANT: Watson, Susan R.
;; APPLICANT: EOS Biotechnology, Inc.
;; TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,
;; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
;; TITLE OF INVENTION: Modulators
;; FILE REFERENCE: 018501-000710US
;; CURRENT APPLICATION NUMBER: US/10/021.660
;; CURRENT FILING DATE: 2001-12-06
;; PRIOR APPLICATION NUMBER: US/09/784,356
;; PRIOR FILING DATE: 2001-02-14
;; PRIOR APPLICATION NUMBER: US 09/637,977
;; PRIOR FILING DATE: 2000-08-11
;; NUMBER OF SEQ ID NOS: 135
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 128
;; LENGTH: 411
;; TYPE: .PRT
;; ORGANISM: Homo sapiens
US-10-021-660-128

Query Match 4.7%; Score 147; DB 14; Length 411;
Best Local Similarity 18.2%; Pred. No. 6.4e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 187 WLLILVIALWFLI-----LDVWSTC-----SOPICAKWTRTEAGSKLS 229
Db 5 WKAVLLALASIAIQYTAITFTAKSPHTCPGLAAGLAERLC-----EESFTAYNLSR 59

Qy 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
Db 60 KTH-----ILLATTRSGSGSFVQLFNQHLDPVLFEP----- 93

Qy 290 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKKKKFKRRE 349
Db 94 -----YHVQNTLIPFTQG-----KSPADRRV 115

Qy 350 SLPEQRSMKGARDRAEYI-----RALRRHLVYVPSRPV-----LS 387
Db 116 MLGASRDLLRSYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGADLV 175

Qy 388 LSSGSWTLKLHFFQEVILGA-----SMRALYIYVRD 416
Db 176 LEEGDCVRCGLNLNLTVAAEACRSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

Qy 417 PRAWIYMLYNKPSLYSLKNVPEHLAKFKIEGKGKCNLNSGYAFEPYELRKLKSKK 476
Db 236 PGILASRSETFRDTY-----RLWRLWYGTGR-----KPNLDVTQL-TTVCSDF 279

Qy 477 SNAVS--LLSHLWLANTAALRINTDLLPTSQYQVAKSEDIHVFPQKTERIFAFIGIPL- 533
Db 280 SNSVSTGLMRPPLKLG-----KYMILVRYVEDLARNPMKTEEYIGYIGIPLD 325

Qy 534 -----SPASLNQILFATSTNLFVLPYEGEISPTNTNWKQNLPRDEIKLIENI 581
Db 326 SHVARNQNTNRGDPITLKGKYGTVRN-----SAATAEKWRFLSLDIVAFAGNA 375

Qy 582 CWTLMRLGY 591
Db 376 CQVLAQLGY 385

RESULT 10
US-10-211-462-97
;; Sequence 97, Application US/10211462

; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-462-97

Query Match 4.7%; Score 147; DB 15; Length 411;
Best Local Similarity 18.2%; Pred. No. 6.4e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY 187 WILILVIALWFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 229
DB 5 WKAVLLALASIAIQTAITFTAKSFHTCPGLAEGLAERLC-----EESPTFAYNLSR 59

QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
DB 60 KTH-----ILILATRSFGSSFGQLFNQHLDFVYLFEPL----- 93

QY 290 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPNRGLAQYFAMNKKRKFQRE 349
DB 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115

QY 350 SLPEQSQMGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 387
DB 116 MLAGSRDLRLSLYDCDYLFIENYKPPVNHHTDRIFRGASRVLCSPVCDPPGPADLV 175

QY 388 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 416
DB 176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

QY 417 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGNLNSGYAFEYELPKELSKSK 476
DB 236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPYNLDTVQL-TTVCEDF 279

QY 477 SNAVS--LLSHLWLANTAAALRINTDLLPTSQVLKFEEDIVHPQKTERIFAFGLIPL- 533
DB 280 SNSVSTGLMRPPLWLG-----KPYNLDTVQL-TTVCEDF 325

QY 534 -----SPASLNQILFATSTNLFYLPYEGEISPTNTVWKNLPRDIKLIENI 581
DB 326 SHVARWQNTNRGDTPLGKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375

QY 582 CWTLMRLGY 591
DB 376 CQOVLAQLGY 385

RESULT 11

US-10-408-765A-395
; Sequence 395, Application US/10408765A
; Publication No. US20040101874A1

; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-395

Query Match 4.7%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 6.4e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY 187 WILILVIALWFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 229
DB 5 WKAVLLALASIAIQTAITFTAKSFHTCPGLAEGLAERLC-----EESPTFAYNLSR 59

QY 230 EGHMDLPDVVITSLPGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 289
DB 60 KTH-----ILILATRSFGSSFGQLFNQHLDFVYLFEPL----- 93

QY 290 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHFNPNRGLAQYFAMNKKRKFQRE 349
DB 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115

QY 350 SLPEQSQMGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 387
DB 116 MLAGSRDLRLSLYDCDYLFIENYKPPVNHHTDRIFRGASRVLCSPVCDPPGPADLV 175

QY 388 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 416
DB 176 LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

QY 417 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKEGGKGNLNSGYAFEYELPKELSKSK 476
DB 236 PRGILASRSETFRDTY-----RLWRLWYGTGR-----KPYNLDTVQL-TTVCEDF 279

QY 477 SNAVS--LLSHLWLANTAAALRINTDLLPTSQVLKFEEDIVHPQKTERIFAFGLIPL- 533
DB 280 SNSVSTGLMRPPLWLG-----KPYNLDTVQL-TTVCEDF 325

QY 534 -----SPASLNQILFATSTNLFYLPYEGEISPTNTVWKNLPRDIKLIENI 581
DB 326 SHVARWQNTNRGDTPLGKHYGTVRN-----SAATAEKWRFLSYDIVAFAQNA 375

QY 582 CWTLMRLGY 591
DB 376 CQOVLAQLGY 385

RESULT 12

US-10-723-860-1544
; Sequence 1544, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01

```

, CURRENT APPLICATION NUMBER: US/10/723,860
,
, CURRENT FILING DATE: 2003-11-26
,
, PRIOR APPLICATION NUMBER: 60/429,739
,
, PRIOR FILING DATE: 2002-11-26
,
, NUMBER OF SEQ ID NOS: 8393
,
, SOFTWARE: PatentIn version 3.2
,
, SEQ ID NO 1544
,
, LENGTH: 411
,
, TYPE: PRT
,
, ORGANISM: Homo sapiens
,
, US-10-723-860-1544

```

Query Match	4.7%	Score 147;	DB 16;	Length 411;	
Best Local Similarity	18.2%	Pred. No. 6.4e-05;			
Matches	89;	Conservative 58;	Mismatches 149;	Indels 194;	Gaps 16;
QY	187	WILLIATLWFIEL-----LDVWSTC-----SQPICAKWTTEAGSKSLSS	229		
Db	5	WKAVLLLASIAIQYTAIRTTAKSFYTCGLAEGLAERLC-----EESPTAYNLSR	59		
QY	230	EGHMDLPDVVITSLPGSGABILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACE	289		
Db	60	KTH-----ILILATTRSGSSFVGQLFNQHLDVDLYLPEPL-----	93		
QY	290	WKVSDIRSGHPRLLRGWLQSLVQDTKYLHQLNHLHEPNRGKLAQYFANVKDKKRFKRE	349		
Db	94	-----YHVQNTLIPRFTQG-----KSPADRRV	115		
QY	350	SLPEORSOMKGAFOFDAEYI-----RALRRHLVYPSARPV-----LS	387		
Db	116	MLGASRDLSRLSYDCDLYFLENYIKPPVNHHTDRIFRGASRVLCSPVCDPPGPADLV	175		
QY	388	LSSGSWTLLKHLFFQVFLGA-----SNRALYIYRD	416		
Db	176	LEEGDCVRKCGLLNLTVAAEACRERSHVAIKTVRVPEVNDLRALVEDPRNLKVQLVRD	235		
QY	417	PRAWIYMLYNSKPSLYSLKNVPEHLAKFKIEGKGKCNLNSGVAFYEPFLRKELSKK	476		
Db	236	PRGILASRSETRDTY-----RLWLWYGTGR-----KPNLVDVTQL-TTVCEDF	279		
QY	477	SNAYS--LLSHLWLANTAAALRINTDLLPTSQVLVKPEDIHVHFPKQTTERTIFAPLGIPL-	533		
Db	280	SNSVSTGLMRPPLWLG-----KYNLVRYEDLARNPKKTEEYIGFLGIFLD	325		
QY	534	-----SPASLNQILFATSNLFYLPYEGEISPTNTNVWKNQLPRDEIKLIENI	581		
Db	326	SHVARWIQNTNRGDDTLGKHXYGTVRN-----SAATAEKRRFRLSYDIVAFAGNA	375		
QY	582	CWTLMDRLGY	591		
Db	376	COQVLAOLGY	385		

```

RESULT 13
US-10-212-933-2
; Sequence 2, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideaki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYL
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA41.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007

```

```

; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-212-933-2

```

	Query Match	4.3%; Score 134; DB 14; Length 483;
	Best Local Similarity	26.2%; Pred. No. 0.0015;
	Matches	53; Conservative 36; Mismatches 81; Indels 32; Gaps 9
Qy	408 MRALYIVRDPRAWTYSMLYNKPSLY--SLKNVPE-----HLAKLFKTEGGK-GKCNLNS	459
	: : : : : : : : : : : :	
Dd	278 LKVILHVRDPPA-VASSRIRSRHGILRESLQVVSRODPRAHRMPPLEAGHKLGAKKEGM	336
	: : : : : : : : : : : :	
Qy	460 GYAFEYEPLRKELSKSNVAVLSSHUWLANTAALRIINTOLLTPSYOLVKFEDYIHPQP	519
	: : : : : : : : : : : : : : : :	
Dd	337 GGPADYHAL-----GAMEVICSNMAKT'LTALQ-PPDWLOGHYLVVRVEDYLVGDPVV	386
	: : : : : : : : : : : : : : : :	
Qy	520 KTTTERIAFLGIPLSPA-----SLNQILLPATSTNLFYLPYEGEISPTN-----TNVMQNQLP	571
	: : : : : : : : : : : : : : : :	
Dd	387 KTLRRVDYFVGLLVSPMEQFALNTSGSSGSSKPFF-----VSARNATQAANWRTALT	441
	: : : : : : : : : : : : : : : :	
Qy	572 RDEIKLIENICWTLMDLRGYPK	593
	: : : : : : : : : : : : : : : :	
Dd	442 FOOIKOVEEFCYOPNAVILGYER	463
	: : : : : : : : : : : : : : : :	

```

RESULT 14
US-10-212-933-4
; Sequence 4, Application US/10212933
; Publication No. US2003008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takaashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLVGLUCOSAMINE-6-0-0
; TITLE OF INVENTION: SULFOTENDRASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-212-933-4

```

	Query Match	4.2%	Score 133;	DB 14;	Length 484;
	Best Local Similarity	26.2%;	Pred. No. 0.0019;		
	Matches 53; Conservative	36;	Mismatches 81;	Indels 32;	Gaps 9
Qy	408 MRALVIVRDPRAWTYSMLYNSKPSLY--SLKNVPE-----HLAKLFKEIEGGR-KGCNLLNS	459			
	: : : : :				
Db	279 LKVTHLVDRDPA-VASSPIRSHGLIRLSLQVRSDPRAHRMPLEAAGHKLGAKKEGV	337			
	: : : : :				
Qy	460 GYAFEYELRKELSKSNVSLSHLWANTAAALRIINTDLLPTSYQLVKFEDIVHPQQ	519			
	: : : : :				

Db 338 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 387
Qy 520 KTERIPAFILGIPLSA-----SLNQILFATSTNLFLPYEGEISPTN----TNVWKQNL P 571
Db 388 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 442
Qy 572 RDEIKLIENICWTIMDRLGYPK 593
Db 443 FQIKQVEEFCYOPMAVLGYER 464

RESULT 15
US-10-723-860-1409
; Sequence 1409, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1409
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-1409

Query Match 4.2%; Score 133; DB 16; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.0021;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 408 MRALYIVRDPRAITYSLYNKPSLY--SLKNYPE-----HLAKLFKIEGK-GKCNLNS 459
Db 325 LKVIHLVRDPRV-VASSRIRESHGLIRESLOVVRSRDPRAHRMPFLEAAGHKLGAKEGV 383
Qy 460 GYAFEYELRKLKLSKSNVSLSHLWLTAAALRINTDLLPTSQVLVKFEDI VHF PQ 519
Db 384 GGPADYHAL-----GAMEVICNSMAKTLQTLQ-PPDWLQGHYLVVRYEDLVGDPV 433
Qy 520 KTERIPAFILGIPLSA-----SLNQILFATSTNLFLPYEGEISPTN----TNVWKQNL P 571
Db 434 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAWRTALT 488
Qy 572 RDEIKLIENICWTIMDRLGYPK 593
Db 489 FQIKQVEEFCYOPMAVLGYER 510

Search completed: June 23, 2005, 09:27:55
Job time : 77.3449 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 13.5183 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-9
Perfect score: 3147
Sequence: 1 MNRNGAMVDVMDAHYKMF.....LIENICWTLMMDRLGYPKFM 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	4.2	484	JE0261	N-acetylglucosamin
2	113.5	3.6	484	JE0261	N-acetylglucosamin
3	109.5	3.5	1441	JE00335	hypothetical prote
4	107	3.4	2401	T28676	rhostry protein -
5	106.5	3.4	532	H96840	hypothetical prote
6	106	3.4	486	JE0261	N-acetylglucosamin
7	105.5	3.4	518	B33472	hypothetical prote
8	103	3.3	582	T46822	phytoene desaturas
9	102.5	3.3	689	S45901	probable membrane
10	102	3.2	1333	S65812	DNA-directed DNA p
11	101.5	3.2	591	B97994	RNA primase (EC 2.
12	101	3.2	486	H81377	exopolysphatase
13	100	3.2	1663	T28923	hypothetical prote
14	98	3.1	480	T20168	hypothetical prote
15	98	3.1	586	H95123	DNA primase (impor
16	97.5	3.1	834	T28250	ORF MSV089 probabl
17	97	3.1	420	A28882	alpha-1-antitrypsi
18	96	3.1	561	G54246	conserved hypothet
19	96	3.1	458	A57397	chondroitin 6-sulf
20	96	3.1	1692	G01449	probable mucin G2
21	96	3.1	4910	S64942	probable membrane
22	95	3.0	522	B86294	hypothetical prote
23	95	3.0	963	C90535	conserved hypothet
24	94	3.0	1405	T27969	hypothetical prote
25	93.5	3.0	804	A39972	segment S5 protein
26	93.5	3.0	825	T32130	hypothetical prote
27	93	3.0	648	S56783	hypothetical prote
28	93	3.0	676	A91349	probable ATP-depen
29	93	3.0	868	S65186	NTP80 protein - ye

conserved hypothet
ABC transport syst
hypothetical prote
hypothetical prote
dynein heavy chain
hypothetical prote
hypothetical prote
probable receptor-
probable 2-acylgly
hypothetical prote
probable CDP glyce
ATP-dependent heli
hypothetical prote
hypothetical prote
limb deformity (ld
hypothetical prote

30 93 3.0 1206 2 F72233
31 92.5 2.9 643 2 H81361
32 92.5 2.9 986 2 D72299
33 92.5 2.9 1070 2 S46755
34 92.5 2.9 4344 1 A53489
35 92 2.9 488 2 T33626
36 92 2.9 860 2 S64366
37 92 2.9 961 2 H84787
38 92 2.9 1170 2 A81368
39 92 2.9 4131 2 T21085
40 91.5 2.9 746 2 S06049
41 91 2.9 862 2 H69107
42 91 2.9 1002 2 AF2383
43 91 2.9 1082 2 T43990
44 91 2.9 1213 2 A41724
45 91 2.9 1245 2 T31927

ALIGNMENTS

RESULT 1

JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C/Species: Homo sapiens (man)
C/Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C/Accession: JE0261
R/Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J. Biochem. 124, 670-678, 1998
A/Title: Human N-Acetylglucosamine-6-O-sulfotransferase Involved in the biosynthesis of (

A/Reference number: JE0261; MUID:98391845; PMID:9722682

A/Accession: JE0261

A/Molecule type: mRNA

A/Residues: 1-484 <UCH>

A/Cross-references: DDBJ:AB014679

C/Comment: This protein catalyzes the transfer of sulfer of sulfate from 3'-phosphoadenos
C/Superfamily: chondroitin 6-sulfotransferase
C/Keywords: sulfotransferase

Query Match	4.2%	Score 133;	DB 2;	Length 484;
Best Local Similarity	26.3%	Pred. No. 0.023;		
Matches	53;	Conservative	36;	Mismatches 81; Indels 32; Gaps 9;
Qy	408	MRALYIVRDPRAVIYMLYNSKPSLY--SLKNVPE-----HLAKLFKEGSK-GKCNLNS	459	
Db	279	LKVIHLVRDPRA-VASSRIRSRHGLIRESLQVVRSDPRRAHRMPFLAAGHLGAKKEGV	337	
Qy	460	GVAPEYEPRLKELSKSNVAVLSLHLWLAATAAALINTDLLPTSQLVKFEIVHFPQ	519	
Db	338	GGPADYHAL-----GAMEVICNSMAKTQLTALQ-PPDWLQGHVLYVRVEDLVGEVP	387	
Qy	520	KTERIFAPFAGIPLSPA-----SLNQILFATSTNLFYLPYEGETSPTN-----TNVWKQLNP	571	
Db	388	KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANARALT	442	
Qy	572	RDEIKLIENICWTLMDBLGYPK	593	
Db	443	FQOIQKQVEEFCYQPMVILGYER	464	

RESULT 2

JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C/Species: Mus musculus (house mouse)
C/Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C/Accession: JC7350
R/Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsuoka, T.; Yamakawa, T.; Kurosawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A/Title: Diversity of N-Acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a r
A/Reference number: JC7350; MUID:20374462; PMID:10913333
A/Accession: JC7350


```

Db      989 LNKIKIIDVKYBTDEIEKNKKNKINAELNSSEKIITQLKENSLSKCEQSKIKSTIDNNY 1048
Qy      366 ---AEVIRALRRHLV-----YPSARPV---LSLSSGSLWTLKHLFFQEVLGSMRA 410
Db      1049 VSECIKMITNLKTVIVKKNINITYFKNABEYNQVSLN-----FNNIEMADTKS 1098
Qy      411 LYIVRDPRAWIYMLYNSKPSLSKKNVPEHLAK--LFKIEGGKG-----KCN----- 456
Db      1099 QYILNKKCN-----NGTNTDYNIKELKEHKKSNNVYKDEAGKNTQBIKKNKELPEKYE 1152
Qy      457 -----LNSGVAFEPYELRKLKSKSNVAVSLSHLWLAN---TAAALRINTDLLPTSQY 507
Db      1153 QEVTVLLNKKYAVE---LKNKFDKTKYSEQIIKEIKDAHNTFTSQADKSEKKNKEIKNE 1209
Qy      508 LKVEEDIVHPQKTERIFAFGLGPLSPASLNQILFATSTNLFYLPVE--GEISPTNTNV 565
Db      1210 QIRIEDEVAKNNKNK---AILDQLSVEFP-KIKFLKIKOLRTKSDCLKETKDIETKI 1265
Qy      566 WKONLPRDEIKLIEN 580
Db      1266 SNLSIDTQETKLIEN 1280

RESULT 5
H96840
hypothetical protein F23A5.i8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96840
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.B.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96840
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-532 <STO>
A:Cross-references: UNIPROT:Q9SAH8; GB:AE005173; NID:G6503294; PIDN:AAF14670.1; GSPDB:GN
C:Genetics:
A:Gene: F23A5.i8
A:Map position: 1
C:Superfamily: natural resistance-associated macrophage protein 1

Query Match 3.4%; Score 106.5; DB 2; Length 532;
Best Local Similarity 19.7%; Pred. No. 2.9;
Matches 125; Conservative 90; Mismatches 210; Indels 209; Gaps 33;

Qy      26 NSPMASIQEAEQ-AAEFKKRTQFNVVTFQMESTITRIAYVYGPY----- 70
Db      23 NSPLIENSDSQIIVSEKSKWNF--FAYLPGFLVLSIAYIDPGNFETDLQAGAHYKEL 80
Qy      71 ---INVSSCRFTDSSNPGLOISLVNNTNTHVSVIVTDYHNLKTRFNLYLFGFGFASVADQG 127
Db      81 LWIILVASC-----AALVTQSLAANLGVVTKHLAE-----QCRAEYS 118
Qy      128 QITRPLGQTQAIKVPVRHD--RII---PPFGKFNIAV--GLIL--CISLVILTFQWRFY 178
Db      119 KVPNFMWVAEIAVACCDIPEVIGTAFALANLFSIPVMIGVLLTGLSTLILLALQKYG 178
Qy      179 LSFRLKMRWILLVIALWFIELLDVWSTCSQPICAKWTTRTEAE--GSKKSLSEGHMDLP 238
Db      179 RKLEFLIAFLVETIAICFFVEL-----HYSPKD 206
Qy      239 -----VVITSLPGSGAEIL-----KQLFFNSSDFLYIRVPTAVIDIPETELE 280

```

```

Db      207 PGEVLHGLFVPQIKGNGATGLAISLIGAMVPHNLFLHLSALVLSRKIPRSASGIKE----- 262
Qy      281 IDSFVDACEKVKSDIRSGHFRLLRGLWQSLVQDTKGLHQLNHLHHPNRGKLAQYFAMNKD 340
Db      263 -----ACREYVL--IESG-----LALMVAFILNVSVISVGAVCNAPNLS----- 299
Qy      341 KKRKFKRRRESLPRQSRQMGAFDRAEYIRALRRHLVYVPSARPVLSLSSGSLWTLKHLFF 400
Db      300 -----PEDRANCE-----DLDLNKASFLLRNVV-----GKWSKUL--F 330
Qy      401 QEVLGASMR-----LYIVR---DPR--AWIYSLMYSKPSLSYSLKKNVPEHLAKLF 446
Db      331 AIALLAGSQSSTTGTGYAGQYVVGQFLDLRLPEWLRNLLTR-----CLAIPSLIVALI 384
Qy      447 KIEGGKCNLNSGYAFEYE-----PLRKELS-KSKS-----NAVSLSHLWLANTAAL 495
Db      385 GGSAGAGKLIIIASMILSPFALVPLLKFTSCKTKGSHVNPMAITALTWTWIG-GLIM 443
Qy      496 RINTDLLPTSQYLVKFEIDIVHFPQKTERIF-----AFLGIPLSASLNQILF---ATSTN 548
Db      444 GINIYVLVSSF--IKL--LIHSHMKLILVVFVCGILSPAGIALYLAAYLVFRKNRVATS 499
Qy      549 LFYLPYEGEISPTNTNVWKQNLPDEIKLIENIC 582
Db      500 LL-----ISRDSQNV--ETLPRQDIVNMQLPC 524

RESULT 6
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7351
R:Uchimura, K.; Sasakihany, F.; Kadamatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: UNIPROT:O75667; DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylatic
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 3.4%; Score 106; DB 2; Length 486;
Best Local Similarity 20.6%; Pred. No. 2.8;
Matches 87; Conservative 45; Mismatches 136; Indels 154; Gaps 18;

Qy      179 LSFRLKMRWILLVIALWFIELLDVWSTCSQPICAKWTTRTEAE--GSKKSLSEGHMDLP 237
Db      187 LTAAALFRWTKVI-----CSPPLCPGAPRAAEVGLVEDTACE---RSCP 230
Qy      238 DWITSLPGSGAEILKQLFFNSSDFLYIRVPTAVIDIPETELEIDSFVDACEWKVSDIRS 297
Db      231 PVAIRALE---AECRK-----YPVVVI-----KQVR- 253
Qy      298 GHFLLRGLWQSLVQDTKGLHQLNHLHHPNRGKLAQYFAMNKDKRFRRESLPEORSQ 357
Db      254 ---LLDLGLVLPFLURDPLGLNKKVQLFRDPRA-----VHNSRLKSRQGLLESRLQLVLR 305
Qy      358 MKGAFDRAEYIRALRRHLVYVPSARPVLSLSSGSLWTLKHLFFQEVLGASMRALYIVRDP 417
Db      306 QRG--DR---FHRVLLAHGV---GARP-----GGQSRALPAA--P 335
Qy      418 RAWIYSLMYSKPSLSYSLKKNVPEHLAKLFKTEGGKCNLNSGYAFEYELRKLKSKS 477
Db      336 RADFF-----LT 342
Qy      478 NAVSLSHLWLANTAALRINTDLLPTSQYLVKFEIDIVHFPQKTERIFAFGLI-----PL 533
Db      343 GALEVICEAWLRDLFLA--RGAPAWLRRRLRYLRVEDLVQRPQRAUQLRRLRFSGLRALAAL 401

```


A;Experimental source: strain S288C

C;Genetics:

A;Gene: MIPS:YBR043C

A;Cross-references: SGD:S0000247

A;Map position: 2R

C;Superfamily: yeast probable membrane protein YBR043C

F;108-124/Domain: transmembrane #status predicted <TM1>

F;140-161/Domain: transmembrane #status predicted <TM2>

F;177-193/Domain: transmembrane #status predicted <TM3>

F;239-256/Domain: transmembrane #status predicted <TM4>

F;285-287/Domain: transmembrane #status predicted <TM5>

F;476-493/Domain: transmembrane #status predicted <TM6>

F;513-529/Domain: transmembrane #status predicted <TM7>

F;555-579/Domain: transmembrane #status predicted <TM8>

F;587-603/Domain: transmembrane #status predicted <TM9>

F;629-645/Domain: transmembrane #status predicted <TM10>

F;648-670/Domain: transmembrane #status predicted <TM11>

Query Match 3.3%; Score 102.5; DB 2; Length 689;

Best Local Similarity 21.8%; Pred. No. 8.6;

Matches 78; Conservative 56; Mismatches 139; Indels 85; Gaps 14;

Qy 167 SLVLTQWRPYLFRKLMRWILVIALWFIEL---DWTSCQPICAK----- 214

Db 253 SLVNRWPRSTQWFMVILSGCNVILLVLPETLRQDSKGAIAQILAERRIQVDNNER 312

Qy 215 -----WTRFAESGK-----KSLSSGCH-----MDLPDVVITSLPG-----SCAEI 251

Db 313 GEIQEDYQRGDETDRIENQVATLSTKHNVYGEVRDQSDLDLSESHSPNTYDGRAGETQ 372

Qy 252 LKQLFNFSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGWLSLV 311

Db 373 LQRIYTEASRLY-----EYQLD-DSGIDATTAQVTRIRSTDPKLARSIRENSL 420

Qy 312 Q--DTKLHLQNLHLEHPRNGKLAQYFAMNKDKRKRRESLPQRSQMGAFDRDAEY- 368

Db 421 RKLQTNLEEQKVLSSNGGEIA-----PRQVSAVRKVMWDTFFVYF 461

Qy 369 IRALRR-HLVYTPGARPVLSLSSGSGSWTLKLFHFFQEVLGASMRALYIVDRPRAWIYMLYN 427

Db 462 IKPLKSLHFLYPPVALAITFSAYS-----FSTVYFVNVTVEYKYSRPP-----YN 507

Qy 428 SKPSLSYSLKNVPEHLAKLF-KIEGKGKCNLSNGYAFYEYPLRKLSKSNVSLLS 484

Db 508 FKPLYIGLLYIPNSVTYFFASIGRWVMDLLKRYKKEKIGILAPEARISWNVTSVIS 565

RESULT 10

S65812 RNA-directed DNA polymerase (EC 2.7.7.49) (clone DW15) - Arabidopsis thaliana retrotrans

N;Alternate names: reverse transcriptase

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Jul-1996 #sequence_revision 27-Feb-1997 #text_change 09-Sep-1997

C;Accession: S65814

R;Wright, D.A.; Ke, N.; Smalle, J.; Hauge, B.; Goodman, H.M.; Voytas, D.F.

submitted to the EMBL Data Library, October 1995

A;Description: Multiple non-LTR retrotransposons in the genome of Arabidopsis thaliana.

A;Reference number: S65811

A;Accession: S65812

A;Molecule type: DNA

A;Residues: 1-1333 <WRI>

A;Cross-references: EMBL:L47193; NID:g976276; PID:g976278

R;Wright, D.A.; Ke, N.; Smalle, J.; Hauge, B.M.; Goodman, H.M.; Voytas, D.F.

Genetics 142, 569-578, 1996

A;Title: Multiple non-LTR retrotransposons in the genome of Arabidopsis thaliana.

A;Reference number: S65813; MUID:97005555; PMID:8852854

A;Accession: S65814

A;Status: nucleic acid sequence not shown

A;Molecule type: DNA

A;Residues: 448-648;714-723;747-757;1086-1103 <WRW>

A;Cross-references: EMBL:L47193

C;Genetics:

A;Mobile element: retrotransposon Tall-1

C;Keywords: nucleotidyltransferase; reverse transcriptase

Query Match 3.2%; Score 102; DB 2; Length 1333;

Best Local Similarity 17.8%; Pred. No. 23;

Matches 100; Conservative 67; Mismatches 156; Indels 238; Gaps 29;

Qy 104 DY--HNLKTRFNVLGGFASVADQGOITRFGTGTAIVPVRHDIRIIFPGFKFNIAVG 161

Db 768 DYIKDLKTRLS---GWFTILSMG-----GKETLLK----- 796

Qy 162 LILCISIVILTFQWRFY-LSFRKLMRWILVIA---LWFIELDVWSTCQPICAKWT 216

Db 797 ---AFALALL-----FYAMSCFKLTTCVNMISMSDFW-----WNALEHKKRKHV 841

Qy 217 RTEAGSKSLSGEGHMDLPVV-----ITSLPGS-GAEILKQLFNFSSDF 262

Db 842 SCE---KNCLSKXGGLGFRDIESFNQALLAKQAWLLQFPNLSLFAFFKSRKYDEBDF 897

Qy 263 LYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRG-----WLQS 309

Db 898 L-----DAELKA---TPSYAWR---SILHGRDLLIKGPRKKVNGSGSTSVWMDP 940

Qy 310 LVQDTKLHLQNLHLEHPRNGKLAQYFAMNKDKK-----RKFKRESLEPQRSQMGKA 361

Db 941 WIYD-----NDP-RLPQKHFSVNLDRVHDLINVEDRCRRDRLEE----- 981

Qy 362 FDRDAEYIRALRRHLVYPS-----ARPVLSLSSGSGSWTLKLFHFFQEVLGASMRALYIV 414

Db 982 -----LFYPADIEIIVKGNPVMSMD----- 1002

Qy 415 RDPRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLSNGYAFYEYPLRKEL-- 472

Db 1003 ----FWW---LHKSXSEYSVK-----SGYWLAFQTNKPELIR 1033

Qy 473 -SKSKSNVSLSHLWLANTAAL-----RINTDLLTSTQLVK-----F 511

Db 1034 EARVQPSNGLKKEIKWSTLSPKILFMRILSSALPVAVQIIRRGMPIDPRQVCGSEG 1093

Qy 512 EDIVH--FPQKTTTIERIAFGIGPLSPASL-NOILFATSTNLFLYPLGEISPTNTNVVKQ 568

Db 1094 ESINHVLFTCSLARQWALSQVPTSQGFQNSIFPANIQYLLELKGKGLIPEQIKKSPW 1153

Qy 569 NLPDEIKLIENICWTLMDRLL 589

Db 1154 VLWR-----LWKVRDKL 1165

RESULT 11

B97994

DNA primase (EC 2.7.7.-) [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004

C;Accession: B97994

R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E.

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M.

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: B97994

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-591 <KUR>

A;Cross-references: UNIPROT:Q8DPW1; GB:AE007317; PID:AAK99782.1; PID:g15458592; GSPDB:G

C;Genetics:

A;Gene: dnaG

C;Superfamily: DNA primase

C;Keywords: nucleotidyltransferase

Query Match 3.2%; Score 101.5; DB 2; Length 591;

Best Local Similarity 19.6%; Pred. No. 8.3;

Qy 577 L---IENICWTLM 587
| | | | |
Db 840 LSHPEHRTKLV 853

RESULT 14
T20168
hypothetical protein C53A5.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20168
R:Mortimore, B.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19232
A:Accession: T20168
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-480 <WIL>
A:Cross-references: UNIPROT:O17700; EMBL:Z81486; PIDN:CAB03989.1; GSPDB:GN00023; CESP:CS53A5
A:Experimental source: clone C53A5
C:Genetics:
A:Gene: CESP:C53A5.9
A:Map position: 5
A:Introns: 28/1; 59/3; 194/1; 264/3; 332/1

Query Match 3.1%; Score 98; DB 2; Length 480;
Best Local Similarity 20.6%; Pred. No. 12;
Matches 94; Conservative 60; Mismatches 162; Indels 140; Gaps 22;

Qy 199 ELLDVWSTCSQPI---CAKWTREAGSKSLSEGHMDLPDVVITSLPGSGAEILKQL 255
| | | | |
Db 67 EVADAWDCQPLEYQDAQFSESEMETQMKEL--EDCSIEASRVFLENLQKS-FKLAKLL 123

Qy 256 ----FFNSDFLYIRVPTAYIDIPETLEIDSFVDACEWKVSDIRSGHFRLLRWGLQSLV 311
| | | | |
Db 124 QLTQSLNPQESSYNR-----KIISTEMK-----KINDTVS-----QLKY 157

Qy 312 QDTKLHLQNLHL-----ENRCKLA---QYFAMNK-----DKKXFKR 347
| | | | |
Db 158 AKRIEMINEHLEKHPGDKGDAFLPNMOPVLIAOQAQAHKPLLGSDPDKTKYK 217

Qy 348 RESLPEQSQMGAFDRAEYIRALRRHLVLYPSARPVLSSSGSWTLKLHFFOE----- 402
| | | | |
Db 218 IGRLPDPRANFAVASSKTNLV-----IGTHNGQFLNKFYYNQKNAR 262

Qy 403 VLGSAMR-----ALYIVRDPRAWIYMLNYSKPSLSL-----KNVPE----- 440
| | | | |
Db 263 CMGANLHKRTKPAAGFHNGALYVVG-----YDSIVLSSVELYDLBEGTWKGFSLNLC 317

Qy 441 -----HLAKLFKEGGKCKNLN-----SGYAFEPYELPKLSKSKSNVSLLSH 485
| | | | |
Db 318 RANAIVACYGEIFVLGGFNGSKNEESIEKTSAGNEFEIFG-EMEGRSFGACVQGR 376

Qy 486 LWLA-----NTAAALRLNDLLPT-----SYOLVKPEDIVHPQKTERIFAFGLIPL 533
| | | | |
Db 377 IYIAGSHSNTNTLKSRSVPDYTKTWDEFSMKNARKAFILH---ATNEAIYAIRGYDE 433

Qy 534 SPASLNQILFATSTNLFY--LPYEGEISPTNTNVWK 567
| | | | |
Db 434 ESALLDQIERDPKCLKWSIVPSKPHVPTSYNYEK 469

RESULT 15
H95123
DNA primase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: H95123
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap
neon, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95123
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-586 <KUR>
A:Cross-references: UNIPROT:Q97QX2; GB:AE005672; PIDN:AAK75185.1; PID:gl4972547; GSPDB:G
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI072

Query Match 3.1%; Score 98; DB 2; Length 586;
Best Local Similarity 18.8%; Pred. No. 15;
Matches 117; Conservative 83; Mismatches 212; Indels 210; Gaps 27;

Qy 87 QISLVNNTTEHVVSIVTDYHNL-KTRNYLGG---FGG-----PASVADQGGITRFLGLQTQ 137
| | | | |
Db 5 QVIEEIKNNANIVTEVIGDVISLQKAGRNYYLQCPFHGKTPSFNVVEDKQFYHCFGCGRS 64

Qy 138 AIVKPVHRHRIIPFGFGFKFNIAVGLILCISLVILTQWRFYLSFRKLMRWLILVIALWPF 197
| | | | |
Db 65 GDV-----PKF-----IEEYQ-----GVPPF 79

Qy 198 IELDVWSTCSQPICAKWTREAGSKSLS---SEGHMDLPD-----VVITS 243
| | | | |
Db 80 IEAVQILG-----QRVGIEVERKPLYSQKSPASPHQALYDMHEDAAKFYHAILMTT 129

Qy 244 LPGSGA-----EILKQLFFN-----SSDFLYIRVPTAY--IDIPETEL-----BID 282
| | | | |
Db 130 TMGEARNYLYQRLTDEVILKHFHGLAPERNYLYQLSDQYREEDLDDSLGLFVLSQAN 189

Qy 283 SFVDACE-----WKVSDIRSGHFRLLRWGLQSLVQDTKLHLQNI 321
| | | | |
Db 190 QFVDTFNRIMPLTNDQGVIAFSGRIMQKTDQTSKYKNSR---STAIENKSYELVYM 246

Qy 322 HLHEPNRKLQAYFAMNKKKPKFRRRESLPEQRSQMKGAPDRD-AEYIRALRRHLVYYP 380
| | | | |
Db 247 DRAKRSSGKASEIYLMGFMVDVIAAYRAGIENAVASMTALSRHVEHLKELTKLV--- 303

Qy 381 SARPVLSLSSGSWTLKLHFFQEVLG--ASMRALYIVRDPRAWIYMLNYSKPSLSLQNV 438
| | | | |
Db 304 -----LVYDGDKAGQAATLKADELIGDMPVQIVSMPPNDLPDEVLQNG 347

Qy 439 PEHLAKLFKEGGKCKNLNSGYAFEPYELPKLSKSKSNVSLLSHLWLANTAAA----- 494
| | | | |
Db 348 PEDLAYLLT---KTRISPIEFYIHOYKPNSENLOAQIEFLEKIAPIVQEKSTAAQNS 403

Qy 495 -LRINTDLLPTSVOLVKFEDIVH-----FPQKTERIFAFGLIPL- 533
| | | | |
Db 404 YIHILADSL-ASPDYTIQIEIVNESRQVRQNRMEGISRPTPIIMPVTVKQLSALIMRAEAH 462

Qy 534 -----SPASLNQIL-----FATSTNLFYLPYE-----GEISP-----TNTNVWKQN 569
| | | | |
Db 463 LLYRMESPLVNDYRLRDEDFAFATPEFVLYDLQGVNLPPEVLAEQTBQTEVERAWYQV 522

Qy 570 LPRD-----EIKLIEN 580
| | | | |
Db 523 LAQDLPAEISPOELSEVENTRN 544

Search completed: June 23, 2005, 08:50:15
Job time : 16.5183 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 65.4049 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-9
Perfect score: 3147
Sequence: 1 MNRNGAMMDVDADHYKMF.....LIENICWTLMDRLGYPKFMD 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3142	99.8	1222	2 Q81ZU8	Q81ZU8 homo sapien
2	1821	57.9	755	2 Q9HAK5	Q9haks homo sapien
3	920	29.2	480	2 Q9H951	Q9h951 homo sapien
4	863	27.4	784	2 Q6P523	Q6p523 homo sapien
5	615	19.5	314	2 Q9HAD7	Q9had7 homo sapien
6	150	4.8	411	2 Q9EQC0	Q9eqc0 mus musculus
7	147	4.7	411	2 Q43916	Q43916 homo sapien
8	136	4.3	958	2 Q9UL01	Q9ul01 homo sapien
9	136	4.3	958	2 Q8BL14	Q8bli4 m mus muscu
10	134	4.3	483	2 Q794G9	Q794g9 mus musculus
11	134	4.3	530	2 Q88276	Q88276 mus musculus
12	134	4.3	530	2 Q80WV3	Q80wv3 mus musculus
13	133	4.2	483	2 Q9U4C5	Q9ued5 homo sapien
14	133	4.2	530	2 Q9V4C5	Q9v4c5 homo sapien
15	129	4.1	304	2 Q677T4	Q677t4 lymphocvsti
16	128	4.1	486	2 Q9VMC3	Q9vmc3 drosophila
17	127	4.0	441	2 Q93403	Q93403 torpedo cal
18	126.5	4.0	420	2 Q6DBY9	Q6dbys brachydanio
19	126	4.0	388	2 Q9WJES	Q9wues mus musculus
20	125	4.0	388	2 Q9RL11	Q9rl1l mus musculus
21	123.5	3.9	395	2 Q9GZX3	Q9gzx3 homo sapien
22	122.5	3.9	384	2 Q6ZE88	Q6ze88 burkholderi
23	122.5	3.9	479	2 Q7LGC8	Q7lgc8 homo sapien
24	122	3.9	392	2 Q6GN39	Q6gn39 xenopus lae
25	122	3.9	515	2 Q8HV31	Q8hv31 renealmia s
26	121.5	3.9	392	2 Q63I21	Q63iz1 burkholderi
27	121	3.8	520	2 Q672W9	Q672w9 curculigo c
28	120.5	3.8	479	2 Q7S099	Q7s099 homo sapien
29	119.5	3.8	390	2 Q7LCN3	Q7lcn3 homo sapien
30	119.5	3.8	411	2 Q9GZS9	Q9gzs9 homo sapien
31	119.5	3.8	518	2 Q9GHE3	Q9ghe3 drymophila

32	119	3.8	257	2 Q794I5	Q794i5 mus musculus
33	119	3.8	472	2 Q88199	Q88199 mus musculus
34	118.5	3.8	358	2 Q7NM85	Q7nm85 chromobacte
35	118	3.7	395	2 Q9QUP4	Q9qup4 mus musculus
36	118	3.7	520	2 Q6LA24	Q6la24 curculigo c
37	117	3.7	515	2 Q8HV32	Q8hv32 renealmia c
38	117	3.7	559	2 Q9FGV3	Q9fgv3 arabidopsis
39	116.5	3.7	120	2 Q9DDK5	Q9ddk5 mus musculus
40	116.5	3.7	363	2 Q9VMC4	Q9vmc4 drosophila
41	116	3.7	315	2 Q8HV29	Q8hv29 renealmia t
42	116	3.7	515	2 Q8HV34	Q8hv34 renealmia a
43	115	3.7	515	2 Q8HV65	Q8hv65 etlingera e
44	114	3.6	515	2 Q8HV21	Q8hv21 siliquomom
45	114	3.6	515	2 Q8HV30	Q8hv30 renealmia s

ALIGNMENTS

RESULT 1
Q81ZU8 ID Q81ZU8 PRELIMINARY; PRT; 1222 AA.
AC Q81ZU8;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE NCAG1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Goossens D., Del-Favero J., Van Broeckhoven C.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480435; AA32895.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer 1; 1.
SQ SEQUENCE 1222 AA; 140272 MW; 156BE397B4E99AE9 CRC64;

Query Match 99.8%; Score 3142; DB 2; Length 1222;

Best Local Similarity 99.8%; Pred. No. 6.1e-243;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MNRNGAMMDVDADHYKMFWDHGHNSPMASIQEAOAAEFKKRWTFQVNVTFQMESTIT	60
Db	627	MNRNGAMMDVDADHYKMFWDHGHNSPMASIQEAOAAEFKKRWTFQVNVTFQMEPTIT	686
Qy	61	RIATVFGPYINSSCRFIDSSNPGLOISLNVNTEHVSVITDYHNLKTRFNYLGGFGF	120
Db	687	RIATVFGPYINSSCRFIDSSNPGLOISLNVNTEHVSVITDYHNLKTRFNYLGGFGF	746
Qy	121	ASVADQGITRFGIGTQAIKVPVHRDRIIPFGPKFNIAGVGLIICISIVILTFQWRPYLS	180
Db	747	ASVADQGITRFGIGTQAIKVPVHRDRIIPFGPKFNIAGVGLIICISIVILTFQWRPYLS	806
Qy	181	FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKKSLSEGHMDLPDVV	240
Db	807	FRKLNRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKKSLSEGHMDLPDVV	866
Qy	241	ITSLPGSGAEILKOLFENSSDFLYIRVPTAYIDIPETELEIDSFVDCAEKWSVDIRSGHF	300
Db	867	ITSLPGSGAEILKOLFENSSDFLYIRVPTAYIDIPETELEIDSFVDCAEKWSVDIRSGHF	926
Qy	301	RLLRGLWQLSVQDTKHLQNLHLHPEHNRGKLAQYFAMNKDKRKFRRRESLPEQRQSKMG	360
Db	927	RLLRGLWQLSVQDTKHLQNLHLHPEHNRGKLAQYFAMNKDKRKFRRRESLPEQRQSKMG	986
Qy	361	AFDRDAEYIRALRRHLVYPSARVLSLSSGSWTLKLHFFQEVILGASMRALYIVRDPRAW	420
Db	987	AFDRDAEYIRALRRHLVYPSARVLSLSSGSWTLKLHFFQEVILGASMRALYIVRDPRAW	1046

```
QY 421 IYSLYNSKPSLSXLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYBPLRKELSKSNAV 480
Db 1047 IYSLYNSKPSLSXLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYBPLRKELSKSNAV 1106
QY 481 SLLSHLWLANTAALRINTDOLLPTSQVLKVFEDIVHPPOKTTTTRIFAFGIPUSPASLNQ 540
Db 1107 SLLSHLWLANTAALRINTDOLLPTSQVLKVFEDIVHPPOKTTTTRIFAFGIPUSPASLNQ 1166
QY 541 ILFATSNLFLVPEGBISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKEMD 596
Db 1167 ILFATSNLFLVPEGBISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKEMD 1222

RESULT 2
QHAKS PRELIMINARY; PRT; 755 AA.
ID QHAKS
AC QHAKS;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ11477.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK021539; BAB13840.1; -.
SQ SEQUENCE 755 AA; 86162 MW; A8104C69FA4D6250 CRC64;

Query Match 57.9%; Score 1821; DB 2; Length 755;
Best Local Similarity 99.4%; Pred. No. 3.7e-137;
Matches 342; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRYNGAMVDVDAHYKMFWDHGHGNSPMASIOEAQAEFKKRWTFQFNVNVTQFQMESTIT 60
Db 411 MNRYNGAMVDVDAHYKMFWDHGHGNSPMASIOEAQAEFKKRWTFQFNVNVTQFQMESTIT 470
QY 61 RIAYVFYGPYINVSSCRFIDSSNPGQLQISLVNNTHTVVSIVTDYHNLKTRFNLYLFGGFG 120
|||||
```

```
Db 471 RIAYVFYGPYINVSSCRFIDSSNPGQLQISLVNNTHTVVSIVTDYHNLKTRFNLYLFGGFG 530
QY 121 ASVADQGQITRGLGTQALVKPVRHDIRIIFPPGFKFNIAVGLICISLVILTFQWRFYLS 180
Db 531 ASVADQGQITRGLGTQALVKPVRHDIRIIFPPGFKFNIAVGLICISLVILTFQWRFYLS 590
QY 181 FRKLNRWILILVIALWFIELDVWSTCSQPICAKWTTRTEAEGSKLSSEGHGMDLDDVV 240
Db 591 FRKLNRWILILVIALWFIELDVWSTCSQPICAKWTTRTEAEGSKLSSEGHGMDLDDVV 650
QY 241 ITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACWKVSDIRSGHF 300
Db 651 ITSLPGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACWKVSDIRSGHF 710
QY 301 RLLRGWLQSLVQDTKHLQNLHLHPENRGKLAQYFAMNKKKK 344
Db 711 RLLRGWLQSLVQDTKHLQNLHLHPENRGKLAQYFAMNKKKK 754

RESULT 3
QH951 PRELIMINARY; PRT; 480 AA.
ID QH951
AC QH951;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ13005.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Kumagai A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK023067; BAB14387.1; -.
SQ SEQUENCE 480 AA; 54559 MW; AEP246851B83B9DB CRC64;

Query Match 29.2%; Score 920; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 4.4e-65;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRYNGAMVDVDAHYKMFWDHGHGNSPMASIOEAQAEFKKRWTFQFNVNVTQFQMESTIT 60
```

```

308 MNRNGAMVDVDAHYKFWDFDHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 367
61 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNVLGFGGF 120
368 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNVLGFGGF 427
121 ASVADQOQITRFGGLGTOAIVKPVVRHDIRIIFPGFKFNIAVGLIICISIVILTF 173
428 ASVADQOQITRFGGLGTOAIVKPVVRHDIRIIFPGFKFNIAVGLIICISIVILTF 480

RESULT 4
ID Q6P523 PRELIMINARY; PRT; 784 AA.
AC Q6P523;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE C18orf4 protein (Fragment).
GN Name=C18orf4;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg K.H., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettner M.C., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.C., Truchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.W., Gough J., Schmutz J., Myers R.M., Burtfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalusz D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RA Strauberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC062557; AAH62557.1; -.
FT NON TER 784 784
SQ SEQUENCE 784 AA; 89666 MW; 2A86FFC4DDCB846C CRC64;

Query Match 27.4%; Score 863; DB 2; Length 784;
Best Local Similarity 99.4%; Pred. No. 3.3e-60;
Matches 161; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRNGAMVDVDAHYKFWDFDHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 60
DB 617 MNRNGAMVDVDAHYKFWDFDHGNSPMASIQEAOAEFKKRWTFQVNVVTFQWESTIT 676
QY 61 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNVLGFGGF 120
DB 677 RIAYVFGPYINVSSCFIDSSNPGLOISLVNNTNTEHVSVITVDYHNLKTRFNVLGFGGF 736
QY 121 ASVADQOQITRFGGLGTOAIVKPVVRHDIRIIFPGFKFNIAVGL 162
DB 737 ASVADQOQITRFGGLGTOAIVKPVVRHDIRIIFPGFKFNIAVGL 778

```

```

RESULT 5
Q9HAD7 PRELIMINARY; PRT; 314 AA.
ID Q9HAD7;
AC Q9HAD7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein FLJ11787.
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Komatsu T.,
RA Togashi T., Oyama K., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK021849; BAB13912.1; -.
FT NON TER 314 314
SQ SEQUENCE 314 AA; 35955 MW; 722B5555DC065538 CRC64;

Query Match 19.5%; Score 615; DB 2; Length 314;
Best Local Similarity 100.0%; Pred. No. 7.6e-41;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 KKSLSSEGHMDLPDVVITSLPGSGAEILKOLFNSDFLYIRVPTAVIDIPETELEIDS 283
DB 198 KKSLSSEGHMDLPDVVITSLPGSGAEILKOLFNSDFLYIRVPTAVIDIPETELEIDS 257
QY 284 FVDACEKWSVDIRSGHFRLLRGLQSLVQDTKLQNLHLHLPNRRGKLAQYFAMNKD 340
DB 258 FVDACEKWSVDIRSGHFRLLRGLQSLVQDTKLQNLHLHLPNRRGKLAQYFAMNKD 314

RESULT 6
Q9EQCO PRELIMINARY; PRT; 411 AA.
ID Q9EQCO;
AC Q9EQCO;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)

```


Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 (6)
 RN RA TISSUE=Brain;
 RP MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RC Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Ustin T.B., Toshlyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN (7)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN (8)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U65637; AAC28776.1; -;
 DR EMBL; AB003791; BAA24840.1; -;
 DR EMBL; AF090137; AAD19878.1; -;
 DR EMBL; AY339617; AAP88041.1; -;
 DR EMBL; BC022567; AAH22567.1; -;
 DR EMBL; BC028235; AAH28235.1; -;
 DR GO; GO:0016021; C:integral to membrane; TAS.
 DR GO; GO:0008146; P:sulfotransferase activity; TAS.
 DR GO; GO:0005976; P:polysaccharide metabolism; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 DR KW Transference.
 SQ SEQUENCE 411 AA; 46714 MW; F07DOA23B6338A09 CRC64;
 Query Match 4.7%; Score 147; DB 2; Length 411;
 Best Local Similarity 18.2%; Pred. No. 0.0042;
 Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;
 187 WILVIALWFIEL-----LDVNSTC-----SOPICAKWTRTEAGSKSLSS 229
 5 WKAVLLALASIAQYTAIRFTAKSFHTCPLAENGLAERLC-----EESPTAYNLRS 59
 230 EGHMDLPDVVITSLPGSGAILKQLFNSSDFYIRVPTAYIDIPETELBIDSFDVACE 289
 60 KTH-----ILILATTRSGSSFGVQLNQHLNDFYLFEPFL----- 93
 290 WKVSDIRSGHFLRLGWLQSLVQDTKLHLNHLHLEPNRGLAQYFANNKDKKFKERE 349
 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115
 350 SLPEQRSMKGAFFORDAEYI-----RALRRHLVYPSARPV-----LS 387
 116 MLAGSRDLRLSLYDCDLFLFENYIKPPVNNHTDIFRGSARVLCSRPVCPDPCPADLV 175
 388 LSSGSWTLLKHLFFQVFLGA-----SNRRLYIVRD 416
 176 LEEGDCVRKCGLLNLTVAAACRERSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

417 PRAMIYMLYNKPSLYSLKQVPEHLAKLPKIEGKGKCNLNSGYAFYEPYPLRKELSKSK 476
 236 PRGLASRSETPRDY-----RLRWLYGTGR-----KPYNLDTQL-TTVCEDF 279
 477 SNAVS--LLSHLMLANTAAALRINTDLLPTSXYLVKKEEDIVHPPOKTERIFAFGLIPL- 533
 280 SNSVSTGLMRPPWLKG-----KYMVLRYVEDLARNPMKTEELIYGLFGLPDL 325
 534 -----SPASLNQILFATSTNLFYLPVEGEISPTNTNVWKNQNLPRDEIKLIENI 581
 326 SHVARMIONNTRGDPITLGKHKYGTVRN-----SAATAEKWRFLSYDIVAFAGNA 375
 582 CWTLMDRLG Y 591
 376 CQOVLAQLGY 385

RESULT 8
 ID Q9UL01 PRELIMINARY; PRT; 958 AA.
 AC Q9UL01; 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
 DE Squamous cell carcinoma antigen recognized by T cell (SART2
 DE protein).
 GN Name=SART-2;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=20143850; PubMed=10679095;
 RA Nakao M., Shichijo S., Imaizumi T., Inoue Y., Matsunaga K., Yamada A.,
 RA Kikuchi M., Tsuda N., Ohta K., Takamori S., Yamana H., Fujita H.,
 RA Itoh K.;
 RA "Identification of a gene coding for a new squamous cell carcinoma
 RT antigen recognized by the CTL";
 RL J. Immunol. 164:2565-2574 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Helel F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Brownstein M.J., Ustin T.B., Toshlyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF098066; AAF00087.1; -;
 DR EMBL; BC039245; AAH39245.1; -;
 DR Genew; HGNC:21144; SART2.
 DR GO; GO:0005783; C:endoplasmic reticulum; TAS.

Thu Jun 23 12:39:06 2005

```
RESULT 10
Q794G9 PRELIMINARY; PRT; 483 AA.
AC Q794G9, 2004 (TrEMBLrel. 27, Created)
DT 05-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-
DE Sulfotransferase-1).
GN Name=Gs6st-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
RW Transferase.
SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FEFD70E CRC64;

Query Match 4.3%; Score 134; DB 2; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.059;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYVRDPRAWIYMLYNKPSLY--SLKNVPE-----HLAKLFKIEGK-KGCNLS 459
DB 278 LKVIHLVRDPA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAAGHKGAKKEGM 336
QY 460 GYAFEYELRKELSKSNVSLSHLWLAALRINTDLPSTYQVLFEDIVHPQ 519
DB 337 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHYLVRYEDLVGDV 386
QY 520 KTERIFAFGLIPSPA----SLNQILPATSTNLFYPYEGEISPTN----TNVWKQNL 571
DB 387 KTLRRVYDFVGLLVSPENEPQALNTSGSSGSKPFV-----VSARNATQANARWALT 441
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 442 FQIQKVEEFCYQPMVLYGER 463

RESULT 11
O88276 PRELIMINARY; PRT; 530 AA.
AC O88276, 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berger J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodríguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
RW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 4.3%; Score 134; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.067;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYVRDPRAWIYMLYNKPSLY--SLKNVPE-----HLAKLFKIEGK-KGCNLS 459
DB 325 LKVIHLVRDPA-VASSRIRSRHGLIRESLQVRSRDPRAHRMPFLEAAGHKGAKKEGM 383
QY 460 GYAFEYELRKELSKSNVSLSHLWLAALRINTDLPSTYQVLFEDIVHPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHYLVRYEDLVGDV 433
QY 520 KTERIFAFGLIPSPA----SLNQILPATSTNLFYPYEGEISPTN----TNVWKQNL 571
DB 434 KTLRRVYDFVGLLVSPENEPQALNTSGSSGSKPFV-----VSARNATQANARWALT 488
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 489 FQIQKVEEFCYQPMVLYGER 510

RESULT 12
O80WV3 PRELIMINARY; PRT; 530 AA.
AC O80WV3, 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Berger J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodríguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
```

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Kzywinski M.I., Skalska U., Smallos D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN
RC SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC051963; AAH51963.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 4.3%; Score 134; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.067;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 325 LKVIHLVRDPRA-VASSRIIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGM 383
QY 460 GYAFEPYELRKLKSKSNVLSHLWLANTAAALRINTDLLPTSQYLVKPEFEDIVHFPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKTLQTALQ-PPDWLQGHYLVVRYEDLVGDPV 433
QY 520 KTTIRIFAFGLIPLSPA-----SLNQILFATSNLFLPYLVEGEISPTN-----TNVWKQNL 571
DB 434 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAAANRALT 488
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 489 FQIKQVEEFCYQPMVGLGYER 510

RESULT 13
QY0UED5 PRELIMINARY; PRT; 483 AA.
AC QY0UED5; 09Y4C5; 09Y6F2; 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.;
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-

RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence.";
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -.
DR EMBL; AB021124; BAB16886.1; -.
DR EMBL; AB021125; BAB16887.1; -.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 4.2%; Score 133; DB 2; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.071;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 278 LKVIHLVRDPRA-VASSRIIRSRHGLIRESLQVRSRDPRAHRMPFLEAGHKLGAKEGV 336
QY 460 GYAFEPYELRKLKSKSNVLSHLWLANTAAALRINTDLLPTSQYLVKPEFEDIVHFPQ 519
DB 337 GGPADYHAL-----GAMEVICNSMAKTLQTALQ-PPDWLQGHYLVVRYEDLVGDPV 386
QY 520 KTTIRIFAFGLIPLSPA-----SLNQILFATSNLFLPYLVEGEISPTN-----TNVWKQNL 571
DB 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAAANRALT 441
QY 572 RDEIKLIENICWTLMRLGYPK 593
DB 442 FQIKQVEEFCYQPMVGLGYER 463

RESULT 14
QY0Y4C5 PRELIMINARY; PRT; 530 AA.
AC QY0Y4C5; 09GZNS; 09Y6F2; 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE (Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawara A., Kadamatsu K., Muramatsu T.;
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells.";
RL J. Biochem. 124:670-678(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=umbilical vein endothelium;
RX MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization.";
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -.
DR EMBL; AF083066; AAD20981.1; -.
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:sulfotransferase activity; IEA.


```
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match      4.2%; Score 133; DB 2; Length 530;
Best Local Similarity 26.2%; Pred. No. 0.08;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 408 MRALYIVRDPRAWIYSLYXSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 459
DB 325 LKVTHLVDRDPA-VASSRIRSRHGLIESQVRSRDPRAHRMPFLAAGHKLGAKEGV 383
QY 460 GYAFEPYELRKLKSKSNVSLSHLWLAALRINTDLLPTSQLVKFEDIVHFPQ 519
DB 384 GGPADYHAL-----GAMEVICNSMAKLTQALQ-PPDWLQGHYLVVRYEDLVGDPV 433
QY 520 KTERIFAFGLIPLSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNWKNQLNP 571
DB 434 KTLRRVYDFVGLLVSPMEQPALNWTSGSGSSKPFV-----VSARNATOAAANWRTALT 488
QY 572 RDEIKLIENICWTLMDRIGYPK 593
DB 489 FQIKQVEEFCYQPMVGLGYER 510

RESULT 15
Q677T4 PRELIMINARY; PRT; 304 AA.
AC Q677T4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Lymphocystis disease virus - isolate China.
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Lymphocystivirus.
OX NCBI_TaxID=256729;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed15194775; DOI=10.1128/JVI.78.13.6982-6994.2004;
RA Zhang Q.Y., Xiao F., Xie J., Li Z.Q., Gui J.F.;
RT "Complete Genome Sequence of Lymphocystis Disease Virus Isolated from China.";
RL J. Virol. 78:6982-6994(2004).
DR EMBL; AY380826; AAU11023.1; -
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 35898 MW; 63C9BDBFB2145C22 CRC64;

Query Match      4.1%; Score 129; DB 2; Length 304;
Best Local Similarity 21.1%; Pred. No. 0.077;
Matches 86; Conservative 62; Mismatches 126; Indels 134; Gaps 22;

QY 188 ILILVIALWFTLELDVWSTSQPCAKWTRTEARGSKSLSEGHMDLPDWITSLPGS 247
DB 8 MLLIILLML-----IW-IFQRPVI-----YEKPKN-----ILLITTRS 43
QY 248 GAETLKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVDSIRSGHFRLLRGWL 307
DB 44 GSSFLGEIFNSRSDVFLFEPL-----WHLNEIYTEYVKVKALF 83
QY 308 Q-SLVQDTKLHLQNH-----LHEPNRGLAQYFAMNKKKKPKKRESLP-----EQR 355
DB 84 NCELVVLKRYLTQNFPPFKRNYSKALCKP--GKTCIY-GLNGDRKYTCDLKQCQPLNDIA 140
QY 356 SOMKGAFOADAEYIRALRHLVYVPSARPVLSLSSGWSMTKLHFFQEVLGASMRALYIVR 415
DB 141 SLYCQTF--DTVVIKTVR-----IRNKTOALEL-----MTQFDIKIHLVR 179
QY 416 DPRAWIYSLYNSKPSLYSLKNVPEHLAKFKIEGKGKCNLNSGYAFEPYELRKLKLS 475
DB 180 DPRG-----SFSNK-----IKTF-----NRDYNF-----KQIAKI 204
```

```
QY 476 KSNVSLLSHLWLAALRINTDLLPTSQLVKFEDIVHFPQKTTTERIPAFGLIPLSP 535
DB 205 QDDIDIYETL-----KDRL--GYLLKYEDLIINPKELPHLFSFCELMFDE 250
QY 536 ASLNLQILFATSTNLFYLPYE-GEISPTNTNVWKNLPRDEIKLIENIC 582
DB 251 EVLKTIDRLTNTNRSG-PYAIGKIKPSG---WKTQLSTFKIDMIENAC 294
```

Search completed: June 23, 2005, 08:49:00
Job time : 67.4049 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 42.6885 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001s:*
- 5: Geneseq2002s:*
- 6: Geneseq2003as:*
- 7: Geneseq2003bs:*
- 8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2056	100.0	395	4	AAY72640 Human gly
2	2056	100.0	395	5	Abb81554 Human cor
3	2056	100.0	395	5	Aae15438 Human dru
4	2056	100.0	395	7	Adi21086 Novel hum
5	2056	100.0	395	8	Adl61235 Human tyr
6	1934	94.1	395	5	Abb81555 Consensus
7	1729.5	84.1	390	5	AAY72639 Human gly
8	1729.5	84.1	390	5	Abb81556 Human int
9	1708	83.1	395	4	AAY72638 Mouse gly
10	1708	83.1	395	5	Aau11275 Murine int
11	1694.5	82.4	418	5	Abb81557 Mouse int
12	1620	78.8	418	3	Aab41947 Human ORF
13	1538	74.8	394	7	Adj70405 Human hea
14	1024	49.8	386	4	Aam93309 Human pol
15	1024	49.8	386	8	Adl30784 Human pro
16	1019	49.6	386	2	AAY39918 Human gly
17	1018	49.5	380	5	Aau11274 Human L-s
18	981	47.7	386	3	AAY79219 Human tra
19	977	47.5	388	2	AAY39919 Mouse gly
20	808	39.3	169	5	Abb81558 Human cor
21	755	36.7	169	5	Abb81559 Human int
22	663	32.2	483	2	AAY31656 Mouse N-a
23	657.5	32.0	530	4	AAB95367 Human pro
24	657.5	32.0	530	8	Adq18590 Human sof
25	653.5	31.8	484	2	AAY31657 Human N-a

26	653.5	31.8	531	5	AAU69414	Lung amal
27	653.5	31.8	531	8	ADRI4283	Human NF-
28	653	31.8	128	7	ADI21571	Novel hum
29	614	29.9	486	6	ABP56121	Human cho
30	578.5	28.1	479	7	ABM85237	Human pro
31	577.5	28.1	499	6	ABR41139	Human DIT
32	576.5	28.0	479	2	AAW52863	Glycosami
33	569	27.7	481	7	ABM85236	Mouse pro
34	555.5	27.0	411	2	AAW61100	Human s
35	555.5	27.0	411	5	AAE25356	Human cho
36	555.5	27.0	411	6	ABU03503	Angiogene
37	555.5	27.0	411	7	ADJ68589	Human hea
38	555.5	27.0	411	8	ADQ18725	Human sof
39	554.5	27.0	410	7	ADN95522	Human BEC
40	546.5	26.6	504	8	ABM85022	Human dia
41	532.5	25.9	458	2	AAW06480	Chick cho
42	521.5	25.4	171	5	ABB81560	Human hig
43	377	18.3	169	5	ABB81561	Human N-a
44	364.5	17.7	174	5	ABB81563	Human cho
45	348	16.9	179	5	ABB81562	Human ker

ALIGNMENTS

RESULT 1

AAAY72640

ID AAY72640 standard; protein; 395 AA.

XX AAY72640;

XX 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta).

KW Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1.

OS Homo sapiens.

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02697, AAD02700.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 4B; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4beta). GST-4 gene is found on chromosome 16q23.1. GST is a type 2

XX membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non

CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
 CC the sulphation activity of GST. GST is also useful in inhibiting a
 CC selectin mediated binding event. GST is useful in gene therapy to treat
 CC disorders such as acute or chronic inflammation, systemic lupus
 CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
 CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, pernicious
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, distress
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX
 SQ Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 4; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.1e-207;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLLSSWRSGSFGQLF 60
 DB 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLLSSWRSGSFGQLF 60
 QY 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 DB 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 QY 121 FQWAVSRALCSPACAPPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
 DB 121 FQWAVSRALCSPACAPPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
 QY 181 LOVLVPLSDPALNRLVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 DB 181 LOVLVPLSDPALNRLVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VVREVCRRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPGLE 300
 DB 241 VVREVCRRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPGLE 300
 QY 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 DB 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 QY 361 VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN 395
 DB 361 VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN 395

RESULT 2
 ABB81554
 ID ABB81554 standard; protein; 395 AA.
 XX
 AC ABB81554;
 XX
 XX
 DT 05-SEP-2002 (first entry)
 XX
 XX Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
 DE
 XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlnAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological; chromosome 16q22.
 XX
 OS Homo sapiens.
 XX
 XX US2002061562-A1.
 XX
 XX 23-MAY-2002.
 XX
 XX 09-AUG-2001; 2001US-00927602.
 XX
 XX 11-AUG-2000; 2000US-00638211.
 PR 11-AUG-2000; 2000US-0325773P.

XX (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX Fukuda MN, Akama TO;
 PI WPI; 2002-507643/54.
 XX N-PSDB; ABB89506.
 DR
 XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
 PT useful for treatment, monitoring and diagnosis of macular corneal
 PT dystrophy.
 XX
 XX Claim 13; Fig 1A-D; 69pp; English.
 PS
 XX The present sequence represents human corneal N-acetylglucosamine-6-
 CC sulfotransferase (I), which is able to catalyze sulfation of keratan
 CC sulfate (KS). Also described is a method for monitoring the effect of
 CC treatments for macular corneal dystrophy (MCD), and detecting
 CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
 CC ophthalmological activity. (I) can be used to treat or prevent macular
 CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
 CC without requiring keratoplasty or keratectomy
 XX
 SQ Sequence 395 AA;
 Query Match 100.0%; Score 2056; DB 5; Length 395;
 Best Local Similarity 100.0%; Pred. No. 3.1e-207;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLLSSWRSGSFGQLF 60
 DB 1 MWLPRVSTAVTALLLAQTLLFLVSRPGSPAGGEARVHVLLSSWRSGSFGQLF 60
 QY 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 DB 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRLVRSVFLCDMDVFDAYLPWRRNLSL 120
 QY 121 FQWAVSRALCSPACAPPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
 DB 121 FQWAVSRALCSPACAPPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
 QY 181 LOVLVPLSDPALNRLVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 DB 181 LOVLVPLSDPALNRLVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VVREVCRRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPGLE 300
 DB 241 VVREVCRRSHVRIAEAAATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPGLE 300
 QY 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 DB 301 AWIHNITHGSGPGARREAFKTSRRNALNVSQAWRHALPFAKIRRVQELCAGALQLLGYRP 360
 QY 361 VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN 395
 DB 361 VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN 395
 RESULT 3
 AAE15438
 ID AAE15438 standard; protein; 395 AA.
 XX
 XX AAE15438;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 XX Human drug metabolising enzyme (DME) -5.
 DE
 XX Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;

KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..35
 FT /label= Signal_peptide
 FT Peptide 1..32
 FT /label= Signal_peptide
 FT Protein 33..395
 FT /note= "Human mature DME-5 protein"
 FT Protein 36..395
 FT /note= "Human mature DME-5 protein"

XX WO200179468-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US011869.

XX 13-APR-2000; 2000US-0197590P.

XX 19-APR-2000; 2000US-0198403P.

XX 28-APR-2000; 2000US-0200185P.

XX 05-MAY-2000; 2000US-0202234P.

XX 11-MAY-2000; 2000US-0203509P.

XX (INCY-) INCYTE GENOMICS INC.

XX Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;

PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;

PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;

PI Au-Young J;

XX WPI; 2002-066363/09.

XX N-PSDB; AAD24670.

XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated with
 PT aberrant expression of DME such as allergy, anemia, asthma, infertility.

XX Claim 1a; Page 131-132; 143pp; English.

XX The invention relates to human drug metabolising enzymes referred as DME
 CC and nucleic acid molecules encoding such enzymes. Polynucleotides of the
 CC invention are useful for assessing toxicity of test compounds and in gene
 CC therapy. Sequences of the invention are useful in the diagnosis,
 CC prevention and treatment of autoimmune/inflammatory disorders such as
 CC acquired immune deficiency syndrome (AIDS), adult respiratory distress
 CC syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune
 CC haemolytic anaemia, contact dermatitis, Crohn's disease,
 CC glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease,
 CC Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis,
 CC osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus,
 CC rheumatoid arthritis, scleroderma, ulcerative colitis, vellois, viral,
 CC bacterial, fungal, parasitic, protozoal, helminthic infections; cell
 CC proliferative disorders such as actinic keratosis, arteriosclerosis,
 CC atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's
 CC syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine
 CC disorders such as disorders of the hypothalamus and pituitary resulting
 CC from lesions such as primary brain tumours, adenomas, infarction
 CC associated with pregnancy, aneurysms, vascular malformations; eye
 CC disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa
 CC ; metabolic disorders such as Addison's disease, cystic fibrosis,
 CC diabetes, goitre, glycogen storage diseases, hypercholesterolaemia,
 CC hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome,
 CC mannosidosis, obesity; gastrointestinal disorders such as dysphagia,
 CC gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental

CC disorders. The present sequence is human DME-5 protein

XX Sequence 395 AA;

XX Query Match 100.0%; Score 2056; DB 5; Length 395;

XX Best Local Similarity 100.0%; Pred. No. 3.1e-207;

XX Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLPRVSSSTAVTALLLAQTFLLLFLVSRPGSPSPAGGEARVHVLVSSWRSFGSFGOLF 60

DB 1 MMLPRVSSSTAVTALLLAQTFLLLFLVSRPGSPSPAGGEARVHVLVSSWRSFGSFGOLF 60

QY 61 NQHPDVFYLMPEAWHVTTLTSSQSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120

DB 61 NQHPDVFYLMPEAWHVTTLTSSQSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120

QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180

DB 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180

QY 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSLREQTAKALARDNGIVLTNGTWEADPGLR 240

DB 181 LQVLYPLSDPALNLRIVHLVRDPRVLRSLREQTAKALARDNGIVLTNGTWEADPGLR 240

QY 241 VREVCRSYHRIAEEATLKPPPLRGYRLVRPEDLAREPLAEIRALYATGLSLTPGLE 300

DB 241 VREVCRSYHRIAEEATLKPPPLRGYRLVRPEDLAREPLAEIRALYATGLSLTPGLE 300

QY 301 AWIHNTHSGGPGARREAFKTSNNALNVSOAQRHALPFAKIRRVQELCAGALQLLYRP 360

DB 301 AWIHNTHSGGPGARREAFKTSNNALNVSOAQRHALPFAKIRRVQELCAGALQLLYRP 360

QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 4

ID ADI21086 standard; protein; 395 AA.

XX ADI21086;

DT 15-APR-2004 (first entry)

XX Novel human protein #61.

XX forensic; nutritional source; damaged tissue; diseased tissue;

KW myeloid cell disorder; lymphoid cell disorder;

KW bone cartilage tissue growth; tendon tissue growth;

KW ligament tissue growth; nerve tissue growth; regeneration; wound healing;

KW tissue repair; tissue replacement; burn; incision; ulcer; cancer; human.

XX Homo sapiens.

XX WO2003025148-A2.

XX 27-MAR-2003.

XX 19-SEP-2002; 2002WO-US029964.

XX 19-SEP-2001; 2001US-0323739P.

XX 13-SEP-2002; 2002US-00323739.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang D;

PI Haley-Vicente D;

XX WPI; 2003-354603/33.

XX N-PSDB; ADI21802.

XX

PT New polynucleotides and secreted proteins, useful for treating myeloid or
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve
PT tissue growth or regeneration, in wound healing, and in tissue repair and
PT replacement.

XX Claim 20; SEQ ID NO 337; 156pp; English.

XX The invention relates to an isolated polynucleotide encoding a
XX polypeptide with biological activity. The polynucleotides and
XX polypeptides are useful in diagnostics, forensics, gene mapping,
XX identification of mutations responsible for genetic disorders and other
XX traits, to assess biodiversity, as nutritional sources or supplements.
XX The polynucleotides may also be used as molecular weight markers.
XX chromosome markers or map related gene positions, or as an antigen to
XX raise anti-DNA antibodies or elicit immune response. The polypeptides are
XX useful for raising antibodies, as markers for tissues in which the
XX corresponding polypeptide is expressed, for re-engineering damaged or
XX diseased tissues, for treating myeloid or lymphoid cell disorders, in
XX bone cartilage, tendon, ligament and/or nerve tissue growth or
XX regeneration, in wound healing, in tissue repair and replacement, in
XX healing of burns, incisions and ulcers, and in treating cancer. The
XX present sequence represents the amino acid sequence of a novel human
XX protein.

XX Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 7; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1e-207;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMLPRVSTAVTALLAQTFLFLVSRPGSPAGGEARVHVLVSSWRSGSFVQGLF 60
DB 1 MMLPRVSTAVTALLAQTFLFLVSRPGSPAGGEARVHVLVSSWRSGSFVQGLF 60
QY 61 NQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
DB 61 NQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
DB 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
QY 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEOTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEOTAKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VVREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
DB 241 VVREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 5
ADL61235

ID ADL61235 standard; protein; 395 AA.

XX ADL61235;

XX 03-JUN-2004 (first entry)

DE Human tyrosine kinase biomarker carbohydrate sulphotransferase 6 protein.
XX predictor set; protein tyrosine kinase; cytostatic; antiangiogenic;
KW vasotrophic; vulnery; pharmacogenomic; drug sensitivity; breast cancer;
KW hypervascular disease; angiogenesis; wound healing scar; human;
KW biomarker; carbohydrate sulphotransferase 6; enzyme.

XX Homo sapiens.
XX WO2004020583-A2.
XX 11-MAR-2004.
XX 26-AUG-2003; 2003WO-US026491.
XX 27-AUG-2002; 2002US-0406385P.
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Huang F, Han X, Reeves KA, Amler L, Fairchild CR, Lee FY;
PI Shaw P;
WI; 2004-239171/22.
N-PSDB; ADL61098.
DR New predictor sets with a plurality of polynucleotides and/or
DR polypeptides whose expression pattern predicts cell response to a
PT compound that modulates protein tyrosine kinase activity, useful in
PT treating breast cancer.
XX Claim 9; SEQ ID NO 159; 649pp; English.

XX The invention relates to a novel predictor set comprising a plurality of
XX polynucleotides and/or polypeptides whose expression pattern is
XX predictive of the response of cells to treatment with a compound that
XX modulates protein tyrosine kinase activity or members of the protein
XX tyrosine kinase pathway. The molecules of the invention demonstrate
XX cytostatic, antiangiogenic, vasotrophic and vulnery activities and may
XX be useful in the field of pharmacogenomics, in particular for determining
XX drug sensitivity and in treating breast cancer, hypervascular diseases,
XX angiogenesis and scars in wound healing. The current sequence is that of
XX a human protein tyrosine kinase biomarker protein of the invention.

XX Sequence 395 AA;

Query Match 100.0%; Score 2056; DB 8; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1e-207;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MMLPRVSTAVTALLAQTFLFLVSRPGSPAGGEARVHVLVSSWRSGSFVQGLF 60
DB 1 MMLPRVSTAVTALLAQTFLFLVSRPGSPAGGEARVHVLVSSWRSGSFVQGLF 60
QY 61 NQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
DB 61 NQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
DB 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
QY 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEOTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEOTAKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VVREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
DB 241 VVREVCRSYHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQLE 300
QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRRVQELCAGALQLLGYRP 360
QY 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLPRLGNGFTWASSTASHPRN 395

RESULT 6

AB081555
ID ABB081555 standard; protein; 395 AA.
XX AC ABB081555;
XX
DT 05-SEP-2002 (first entry)
XX
DE Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Homo sapiens.
OS Mus musculus.
OS Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 10
FT /label= Ala, Thr, Val
FT Misc-difference 13
FT /label= Ala, Val, Ser
FT Misc-difference 20
FT /label= Phe, Cys, Gly
FT Misc-difference 39
FT /label= Ala, Asp, Glu
FT Misc-difference 96
FT /label= Val, Met, Ile
FT Misc-difference 142
FT /label= Ala, Thr, Asn
FT Misc-difference 147
FT /label= Ala, Asp, Glu
FT Misc-difference 159
FT /label= Thr, Ser, Gly
FT Misc-difference 238
FT /label= Gly, His, Arg
FT Misc-difference 294
FT /label= Ser, Thr, Gly
FT Misc-difference 371
FT /label= Ala, Thr, Ser
FT Misc-difference 380
FT /label= Leu, Pro, Met
FT Misc-difference 382
FT /label= Gly, His, Ser
FT Misc-difference 384
FT /label= Thr, Ser, Lys
FT Misc-difference 390
FT /label= Ala, Glu
FT Misc-difference 391
FT /label= Ser, Lys
FT Misc-difference 392
FT /label= His, Gln
FT Misc-difference 394
FT /label= Arg, Glu
FT Misc-difference 395
FT /label= Asn, Ser
XX
XX US2002061562-A1.
XX
XX 23-MAY-2002.
XX
XX 09-AUG-2001; 2001US-00927602.
XX
XX 11-AUG-2000; 2000US-00638211.
XX 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
XX (AKAWA/) AKAWA T O.
XX
XX Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
XX

PT New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX PS Example 5; Fig 2A-B; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratoplasty or keratectomy. The present sequence
CC represents a consensus N-acetylglucosamine-6-sulfotransferase which is
CC given in the exemplification of the present invention
XX
XX SQ Sequence 395 AA;
Query Match 94.1%; Score 1934; DB 5; Length 395;
Best Local Similarity 94.4%; Pred. No. 2.2e-194;
Matches 371; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
QY 1 MWLPRVSTAVTALLLAQTFLLLPLVSRPGSPSPAGGEARVHVLVSSWSGSGSVGQLF 60
DB 1 MWLPRFSSTXTVXLLLAQTLLPLVSRPGSPSPAGGEXRHHVLVSSWSGSGSVGQLF 60
QY 61 NQHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
DB 61 SQHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
QY 121 FQAVSRALCSPSPACSAFPRGAISSBAVCKPLCARQSFLLAREACRSYSHVVLKEVRFFN 180
DB 121 FQAVSRALCSPSPACSAFPRGXISSEXVKPLCARQFPFLAREACRSYSHVVLKEVRFFN 180
QY 181 LQVLYPLSDPALNLRVHVRDPRVLRPRVLRPRVLRPRVLRPRVLRPRVLRPRVLRPRVLR 240
DB 181 LQVLYPLSDPALNLRVHVRDPRVLRPRVLRPRVLRPRVLRPRVLRPRVLRPRVLRPRVLR 240
QY 241 VREVCRSVRIAEAAATLKPPLRGYRLVRVEDLAREPDLAREPDLAREPDLAREPDLAREPDL 300
DB 241 VREVCRSVRIAEAAATLKPPLRGYRLVRVEDLAREPDLAREPDLAREPDLAREPDLAREPDL 300
QY 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAVRHALPFAKIRRVQELCAGALQLLYRYP 360
DB 301 AWIHNTHGSGPGARREAFKTSRNALNVSQAVRHALPFAKIRRVQELCAGALQLLYRYP 360
QY 361 VYSEDEQRDLXLDLVLPRGXDXFXWASSTXXXP 393
DB 361 VYSEDEQRDLXLDLVLPRGXDXFXWASSTXXXP 393
RESULT 7
AAY72639
ID AAY72639 standard; protein; 390 AA.
XX
XX AC AAY72639;
XX
XX 02-MAY-2001 (first entry)
XX
XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
XX
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1.
XX

OS Homo sapiens.
PN WO200106015-A1.
XX 25-JAN-2001.
XX 19-JUL-2000; 2000WO-US019741.
XX 20-JUL-1999; 99US-0144694P.
PR 13-JUN-2000; 2000US-00593828.
XX (REGC) UNIV CALIFORNIA.
PA
XX
XX Rosen SD, Lee JK, Hemmerich S;
PI WPI; 2001-138471/14.
DR N-PSDB; AAD02697, AAD02699, AAD02699.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT diagnostic and therapeutic agent screening applications.
PT
XX
XX Claim 3; Fig 1; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
CC alpha). GST-4 gene is found on chromosome 16q23.1. GST is a type 2
CC membrane protein useful for inhibiting a binding event between a selectin
CC and a selectin ligand, which comprises contacting the selectin with a non
CC -sulphated selectin ligand, GST and a small molecular agent that inhibits
CC the sulphation activity of GST. GST is also useful in inhibiting a
CC selectin mediated binding event. GST is useful in gene therapy to treat
CC disorders such as acute or chronic inflammation, systemic lupus
CC erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa,
CC polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation. Note: The present sequence is also shown in
CC sequence listing (page no: 56) but lacks four nucleotides at its 3' end
XX
SQ Sequence 390 AA;
Query Match 84.1%; Score 1729.5; DB 4; Length 390;
Best Local Similarity 85.8%; Pred. No. 7.5e-173; Mismatches 17; Gaps 1;
Matches 333; Conservative 17; Indels 1; Gaps 1;
QY 1 MWLPRVSVSTAVTALLAQ-TFLLFLVSRPGPSPAGGEARVHVLVLSWRSGSFVGLQ 59
DB 1 MWLPRSSKTVTVLLLAQTTCLLFIISRPGPSPAGGEDRVHVLVLSWRSGSFVGLQ 60
QY 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLS 119
DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYMPQSNLSA 120
QY 120 LFQWAVSRALCSPACSPAGGAPRGASISSEAVCKPLCARQSFLLAREACRSYSHVVVKVRRFF 179
DB 121 FFWNAVSRALCSPACSPAGGAPRGITISKQDVCKTCTRQPFSLAREACRSYSHVVVKVRRFF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSDREQTAKALARDNGIVLGTNGTWEADPGL 239
DB 181 NLQVLYPLSDPALNLRIVHLVRDPAVLRSDREAGPILARONGIVLGTNGKWEADPHL 240
QY 240 RVVRECRSHVRAEATLKPFPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPQL 299
DB 241 RLIREVCRSHVRAEATLKPFPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPQL 300
QY 300 EAMHNITHGSGPGARREAPKTSRRNALNVQAWRHLPKARRVOELCAGALQLLYGR 359
DB 301 EAMHNITHGSGGKPEIAPHTSSRNARNVQAWRHLPFTKILRVQEVACAGALQLLYGR 360
QY 360 PVYSEDEQRNALDLVLPRGLNGFTWAS 387

DB 361 PVYSADQQRDLTLDLVLRGPDHFSWAS 388
RESULT 8.
ID ABB81556
XX ABB81556 standard; protein; 390 AA.
AC ABB81556;
XX 05-SEP-2002 (first entry)
XX Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX Homo sapiens.
XX US2002061562-A1.
XX 23-MAY-2002.
XX 09-AUG-2001; 2001US-00927602.
XX 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
PI Fukuda MN, Akama TO;
XX WPI; 2002-507643/54.
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
PT useful for treatment, monitoring and diagnosis of macular corneal
PT dystrophy.
XX Example 5; Fig 2A-B; 69pp; English.
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyse sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratinoplasty or keratectomy. The present sequence
CC represents human intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 390 AA;
Query Match 84.1%; Score 1729.5; DB 5; Length 390;
Best Local Similarity 85.8%; Pred. No. 7.5e-173; Mismatches 17; Gaps 1;
Matches 333; Conservative 17; Indels 1; Gaps 1;
QY 1 MWLPRVSVSTAVTALLAQ-TFLLFLVSRPGPSPAGGEARVHVLVLSWRSGSFVGLQ 59
DB 1 MWLPRSSKTVTVLLLAQTTCLLFIISRPGPSPAGGEDRVHVLVLSWRSGSFVGLQ 60
QY 60 FNOHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLS 119
DB 61 FSQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYMPQSNLSA 120
QY 120 LFQWAVSRALCSPACSPAGGAPRGASISSEAVCKPLCARQSFLLAREACRSYSHVVVKVRRFF 179
DB 121 FFWNAVSRALCSPACSPAGGAPRGITISKQDVCKTCTRQPFSLAREACRSYSHVVVKVRRFF 180
QY 180 NLQVLYPLSDPALNLRIVHLVRDPAVLRSDREQTAKALARDNGIVLGTNGTWEADPGL 239

DR N-PSDB; AAS16948.
XX New enzyme, useful for modifying acceptor molecule, comprises an isolated
PT L-selectin sulfotransferase-2 that directs expression of L-selectin
PT ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal
PT GlcNAc 6-sulfotransferase.
XX
PS Claim 28; Fig 10; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated beta1,3-N-
CC acetylglucosaminyltransferase (beta1,3Gnt) or an active fragment, where
CC beta1,3Gnt directs expression of a MECA-79 antigen. The invention also
CC provides a method of treating or preventing an L-selectin-mediated
CC condition by reducing the expression or activity of a beta1,3Gnt that
CC directs expression of a MECA-79 antigen. This can be done by
CC administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3Gnt,
CC and/or a beta1,3Gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3Gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST
XX
SQ Sequence 395 AA;

Query Match 83.1%; Score 1708; DB 5; Length 395;
Best Local Similarity 83.0%; Pred. No. 1.4e-170;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSRRSGSSFVGOLF 60
DB 1 MRLPRFSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVLSRRSGSSFVGOLF 60
QY 61 NQHPDVFYLMPEAHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSL 120
DB 61 SQHPDVFYLMPEAHVMTTLLSQSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNLSL 120
QY 121 FQWAVSRALCSPPCSAFPRGAISSAEVCKPLCARQSFLLAREACSSYSHVVLKEVRFFN 180
DB 121 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFFN 180
QY 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSRQTKALARDNGIVLGTNGTWVEADPRLR 240
QY 241 VVRECRSHVRIAEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLSLTPQLE 300
DB 241 VVNEVCRSHVRIAEALHKKPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLSLTPQ 300
QY 301 AWTIHNTGSGPGARREAFKTSRRNALNVSOAWRHLPFAKIRRVQELCAGALQLLGYRP 360
DB 301 TWIHNTGSGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 360
QY 361 VYSEDEQRNLALDLVLPRLGNGFTWASSTASHPRN 395
DB 361 VHSELRQDLSLDLLPRGNDSEFKWASSTEKQPE 395

RESULT 11

ABB81557

ID ABB81557 standard; protein; 418 AA.

XX

AC ABB81557;

XX

DT 05-SEP-2002 (first entry)

XX

DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
XX
KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Mus musculus.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PF 09-AUG-2001; 2001US-00927602.
XX
PR 11-AUG-2000; 2000US-00638211.
PR 11-AUG-2000; 2000US-0325773P.
XX
XX (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
XX WPI; 2002-507643/54.
DR
XX New nucleic acid encoding corneal N-acetylglucosamine-6-sulfotransferase,
XX useful for treatment, monitoring and diagnosis of macular corneal
XX dystrophy.
XX
XX Example 5; Page 24-25; 69pp; English.
XX
XX The present invention describes human corneal N-acetylglucosamine-6-
CC sulfotransferase (I), which is able to catalyze sulfation of keratan
CC sulfate (KS). Also described is a method for monitoring the effect of
CC treatments for macular corneal dystrophy (MCD), and detecting
CC susceptibility to MCD. (I) is located to chromosome 16q22, and has
CC ophthalmological activity. (I) can be used to treat or prevent macular
CC corneal dystrophy types I or II. (I) makes possible treatment of MCD
CC without requiring keratinoplasty or keratectomy. The present sequence
CC represents mouse intestinal N-acetylglucosamine-6- sulfotransferase,
CC which is given in comparison with (I) in the exemplification of the
CC present invention
XX
SQ Sequence 418 AA;
Query Match 82.4%; Score 1694.5; DB 5; Length 418;
Best Local Similarity 82.8%; Pred. No. 4.1e-169;
Matches 327; Conservative 24; Mismatches 43; Indels 1; Gaps 1;
QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPAGGEARVHVLVLSRRSGSSFVGOLF 60
DB 25 MRLPRFSTVMSLLMVQTGILVFLVSRQVPSPAGLGERVHVLVLSRRSGSSFVGOLF 84
QY 61 NQHPDVFYLMPEAHVMTTLLSQSAAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSL 120
DB 85 SQHPDVFYLMPEAHVMTTLLSQSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRRNLSL 144
QY 121 FQWAVSRALCSPPCSAFPRGAISSAEVCKPLCARQSFLLAREACSSYSHVVLKEVRFFN 180
DB 145 FQWAVSRALCSPVCEAFARGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFFN 204
QY 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSRQTKALARDNGIVLGTNGTWVEADPGLR 240
DB 205 LQVLYPLSDPALNLRIVHLVRDPRAVLRSRQTKALARDNGIVLGTNGTWVEADPRLR 264
QY 241 VVRECRSHVRIAEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYFTGLSLTPQLE 300
DB 265 VVNEVCRSHVRIAEALH- KPPPLQDRYRLVRYEDLARDPLTVIRELYAFTGLSLTPQ 323
QY 301 AWTIHNTGSGPGARREAFKTSRRNALNVSOAWRHLPFAKIRRVQELCAGALQLLGYRP 360
DB 324 TWIHNTGSGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCGGALQLLGYRS 383
QY 361 VYSEDEQRNLALDLVLPRLGNGFTWASSTASHPRN 395

Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM; Warnock DE;

Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM; Warnock DE;

XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 XX Claim 1; SEQ ID NO 2211; 180pp; English.
 XX
 CC This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytoskeletal activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 XX Sequence 394 AA;

Query Match 74.8%; Score 1538; DB 7; Length 394;
 Best Local Similarity 93.3%; Pred. No. 1.1e-152;
 Matches 294; Conservative 7; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPGAGGEARVHVLVSSWRSGSFVQOLF 60
 DB 1 MWLPRVSTAVTALLAQTLLFLVSRPGSPGAGGEARVHVLVSSWRSGSFVQOLF 60
 QY 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
 DB 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLSDL 120
 QY 121 FQAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
 DB 121 FQAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
 QY 181 LQVLYPLSDPALNLRIVHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 DB 181 LQVLYPLSDPALNLRIVHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGLR 240
 QY 241 VREVCRSYHRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAIRALYFTGLTLPQLE 300
 DB 241 LIREVCRSYHRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAIRALYFTGLTLPQLE 300
 QY 301 AWIHNTHGSGPGAR 315
 DB 301 AWIHNTHGSGPGAR 315

RESULT 14
 AAM93309
 ID AAM93309 standard; protein; 386 AA.
 XX
 AC AAM93309;
 XX
 XX 06-NOV-2001 (first entry)
 DT Human polypeptide, SEQ ID NO: 2817.
 XX Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX Homo sapiens.
 OS
 XX
 XX EPI130094-A2.
 XX

PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94229.
 XX
 XX 930 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 PS Claim 8; SEQ ID NO 2817; 1380pp + Sequence Listing; English.
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA
 CC libraries that were synthesised by the oligo-capping method. The primers
 CC enable the production of the full length cDNA easily without any special
 CC methods. The present sequence is a polypeptide encoded by a full length
 CC human cDNA of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in CD-ROM
 CC format directly from EPO
 XX
 XX Sequence 386 AA;

Query Match 49.8%; Score 1024; DB 4; Length 386;
 Best Local Similarity 54.8%; Pred. No. 1.6e-98;
 Matches 217; Conservative 44; Mismatches 107; Indels 28; Gaps 7;
 QY 14 LLLAQTFLLLFLVSRP-----GSSSPAGGEARVHVLVSSWRSGSFVQOL 59
 DB 2 LLLPKMKLLFLVSRQAILALFFHMYSHNISLSMKAKQPERMHVLVSSWRSGSFVQOL 61
 QY 60 FQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLS 118
 DB 62 FQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRRNLS 121
 QY 119 DLFWQAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 178
 DB 122 SLFWQAVSRALCSPACSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 181
 QY 179 FNLQVLYPLSDPALNLRIVHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADP 237
 DB 182 FNLQVLYPLSDPALNLRIVHVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADP 241
 QY 238 GLRVVREVCRSYHRIAEAAATLKPPFLRGYRLVRFPEDLAREPLAIRALYFTGLTSLT 296
 DB 242 PYYVMQVICQSQLEIYK--TIQSLPKALQERYLLVRVEDLAREPVAQTSRRMFEVLEFL 299
 QY 297 POLEAWIHNTHGSGPGARREAFKTSRRNALVNSQAWRHALPFKIRRVQELCAGALQLL 356
 DB 300 PHLQVTHVNIITRGKMG--DHAFHTNARDALNVQAWRSLPYEKVSRLOKACGDAMNLL 357
 QY 357 GYRHYRSEDEQRNALDLVLPRLGNGFTWASSTASH 392
 DB 358 GYRHYRSEDEQRNALDLVLPRLGNGFTWASSTASH 392

RESULT 15
 ADL30784
 ID ADL30784 standard; protein; 386 AA.
 XX

AC ADL30784;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human protein encoded by a full length cDNA clone SeqID 2817.
XX
KW human; medicine; signal transduction; glycoprotein; transcription;
XX oligo-capping method.
XX
OS Homo sapiens.
XX
PN EF1396543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2004-204755/20.
DR N-PSDB; ADL30783.
XX
XX New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
PS Example 1; SEQ ID NO 2817; 1340pp; English.
XX
XX This invention relates to a novel primers useful for synthesizing full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
XX length human protein of the invention.
SQ Sequence 386 AA;

Query Match 49.8%; Score 1024; DB 8; Length 386;
Best Local Similarity 54.8%; Pred. No. 1.6e-98;
Matches 217; Conservative 44; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTFLLLFLVSRP-----GPSPGAGEARVHVVLSSWRSGSFVQQL 59
Db 2 LLPKMKLLFLVSSQMAILALFFHMYGHNISSLSMKQAQPERMHVVLSSWRSGSFVQQL 61
Qy 60 FNOHPDVFYLMPEPAWHVYNTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRENLS 118
Db 62 FGQHPDVFYLMPEPAWHVYNTLTKQTAMWMLHMAVRDLIRAVFLCDMSVFDAYMECPRRQS 121
Qy 119 DLQWAVSRALCSPAPCPGASISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 178
Db 122 SLQWENSRALCSPACDIIIPQDEIIIPRAHCRLLCSQPPFEVVEKACRSYSHVVLKEVRF 181
Qy 179 FNLQVLPFLSDPALNIRIVHLVRDPRAVLRSRQTAKALARDNGIVLGTNGTWV-EADP 237
Db 182 FNLQSLYPLLDPSLNLHIVHLVRDPRAVFRSRETRKGLMIDNRIVMGQHEQKLKEDQ 241
Qy 238 GLRVVRECRSHVRIAEATLKP-PFLRGYRLVRFDLAREPLAEIRALYAFGLSLT 296
Db 242 PYYVMQICQSLQEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRWYFVGLLEFL 299
Qy 297 POLBAWIHNITHGSGPGGARREAFKTSRNALNVSQAWRHALPPAKIRRVQELCAGALQLL 356

Db 300 PHLQTVVHNITRGKMG--DHAFHTNARDALNVSQAWRWSLPEYKYSRLQKACGDAMNLL 357
Qy 357 GYRPVYSEDEQRNLALDLVLPRGLNGFTWASSTASH 392
Db 358 GYRHVSEQEQRNLLDLL-----STWTVPEQIH 386

Search completed: June 23, 2005, 08:43:27
Job time : 44.6885 secs

THIS PAGE BLANK (USPTO)

GanCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 14.8883 Seconds
(without alignments)
1980.512 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSTAVTALLAQTFF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:**
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2056	100.0	395	4	US-09-949-016-7011 Sequence 7011, Ap
2	1729.5	84.1	390	4	US-09-949-016-6813 Sequence 6813, Ap
3	1729.5	84.1	431	4	US-09-949-016-8893 Sequence 8893, Ap
4	1019	49.6	386	3	US-09-045-284A-2 Sequence 2, Appli
5	1019	49.6	386	3	US-09-130-911-1 Sequence 1, Appli
6	981	47.7	386	4	US-09-786-240-11 Sequence 11, Appl
7	663	32.2	483	3	US-09-263-023-2 Sequence 2, Appli
8	663	32.2	483	4	US-09-471-867-2 Sequence 2, Appli
9	657.5	32.0	608	4	US-09-949-016-9449 Sequence 9449, Ap
10	653.5	31.8	484	3	US-09-263-023-4 Sequence 4, Appli
11	653.5	31.8	484	4	US-09-471-867-4 Sequence 4, Appli
12	653.5	31.8	531	4	US-09-949-016-6471 Sequence 6471, Ap
13	576.5	28.0	479	2	US-08-899-514-2 Sequence 2, Appli
14	555.5	27.0	411	3	US-09-015-188-2 Sequence 2, Appli
15	532.5	25.9	458	2	US-08-655-878-2 Sequence 2, Appli
16	100.5	4.9	199	4	US-09-252-991A-28497 Sequence 28497, A
17	96	4.7	566	2	US-08-484-931B-41 Sequence 41, Appl
18	96	4.7	566	2	US-08-484-158B-41 Sequence 41, Appl
19	96	4.7	566	2	US-08-484-596A-41 Sequence 41, Appl
20	96	4.7	566	2	US-08-480-150A-41 Sequence 41, Appl
21	96	4.7	566	3	US-08-458-731-41 Sequence 41, Appl
22	96	4.7	566	3	US-08-149-223A-41 Sequence 17, Appl
23	95.5	4.6	268	4	US-09-818-780-17 Sequence 94, Appl
24	95.5	4.6	268	4	US-09-818-780-94 Sequence 32986, A
25	93.5	4.5	575	4	US-09-252-991A-32986 Sequence 7, Appli
26	93.5	4.5	1385	2	US-08-687-399-7 Sequence 11, Appl
27	93	4.5	359	3	US-09-150-133-11

ALIGNMENTS

RESULT 1

US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7011

Query Match 100.0%; Score 2056; DB 4; Length 395;
Best Local Similarity 100.0%; Pred. No. 7.5e-210;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MWLPVSTAVTALLAQTFFLLFLVSRPGSPSSPAGGEARVHVLLVSSWRSRSSFGVGLF	60
Db	1	MWLPVSTAVTALLAQTFFLLFLVSRPGSPSSPAGGEARVHVLLVSSWRSRSSFGVGLF	60
Qy	61	NQHPDVFYLMPEAHVWVTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL	120
Db	61	NQHPDVFYLMPEAHVWVTTLSQGSAAATHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL	120
Qy	121	FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN	180
Db	121	FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN	180
Qy	181	LQVLYPLSDPALNRLRVHVRDPRVLRSEQTAKALARDNGIVLCTNGTWTWEADPGLR	240
Db	181	LQVLYPLSDPALNRLRVHVRDPRVLRSEQTAKALARDNGIVLCTNGTWTWEADPGLR	240
Qy	241	VVREVCCHVRIAEAAATLKPPPFLLRGYRLVRFPEDLAREPLAEIRALYAFVGLSLTPGLE	300
Db	241	VVREVCCHVRIAEAAATLKPPPFLLRGYRLVRFPEDLAREPLAEIRALYAFVGLSLTPGLE	300
Qy	301	AWIHNTHGSGPGARREAFKTSRRNALNVSOAQRHALPFKIRRVQELCAGALQLLYRYP	360

28	93	4.5	359	3	US-09-150-141-11	Sequence 11, Appl
29	93	4.5	359	3	US-09-374-493-11	Sequence 11, Appl
30	93	4.5	359	3	US-09-374-824-11	Sequence 11, Appl
31	93	4.5	359	3	US-09-374-492-11	Sequence 11, Appl
32	93	4.5	359	4	US-09-785-343-11	Sequence 11, Appl
33	93	4.5	359	4	US-10-411-976-11	Sequence 11, Appl
34	92	4.5	1093	4	US-09-252-991A-21827	Sequence 21827, A
35	91.5	4.5	557	3	US-09-413-814-34	Sequence 34, Appl
36	91	4.4	421	4	US-09-902-540-10244	Sequence 10244, A
37	89.5	4.4	361	4	US-09-902-540-12804	Sequence 12804, A
38	89.5	4.4	449	4	US-09-984-880-2	Sequence 2, Appli
39	89.5	4.4	449	4	US-10-277-032-2	Sequence 2, Appli
40	88.5	4.3	375	4	US-09-489-039A-11560	Sequence 11560, A
41	88.5	4.3	485	4	US-09-252-991A-31157	Sequence 31157, A
42	88.5	4.3	714	4	US-09-252-991A-29806	Sequence 29806, A
43	88.5	4.3	2388	4	US-09-695-795A-2	Sequence 2, Appli
44	88.5	4.3	6095	3	US-09-144-085-2	Sequence 2, Appli
45	87.5	4.3	377	3	US-09-150-133-5	Sequence 5, Appli

```

Db      301  ATWHTNTHGSGPGARREAFKTSRNALNVQAWRHALPFAKIRRVQELCAGALQLLGYR  360
Qy      361  VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN  395
Db      361  VYSEDEQRNLALDLVLPRLNGFTWASSTASHPRN  395

RESULT 2
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6813

Query Match      84.1%; Score 1729.5; DB 4; Length 390;
Best Local Similarity 85.8%; Pred. No. 3.9e-175;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy      1  MWLPRVSGTAVTALLAQ-TPELLFLVSRPGSPGAGEARVHVLVLSWSRSGSFVQQL  59
Db      1  MWLPRFSSKTVTVLLLAQTTCLLJFIISRPGSPGAGEDRVHVLVLSWSRSGSFVQQL  60

Qy      60  FNQHPDVFYLMPEPAHWVWTTLSQGSAA TLHMAVRDLVRSVFLCMDVPDAYLPWRRNLSD  119
Db      61  FSQHPDVFYLMPEPAHWVWTTLSQGSAA TLHMAVRDLMRSIFLCMDVDFDAYMPQSRNLSA  120

Qy      120  LFWAVSRALCSPPACSAFFPGATISSEAVCKPLCARQSFTLAREACRSYSHVWLKEVRF  179
Db      121  FFWAVSRALCSPPACSAFFPGATISQDVCKTLCTROPFSLAREACRSYSHVWLKEVRF  180

Qy      180  NLQVYLLSDPALNLRIVHLVRDPRAVLRREQOTAKALARDNGIVLGTNGTWVEADPGL  239
Db      181  NLQVYLLSDPALNLRIVHLVRDPRAVLRREAGPILARDNGIVLGTNGKWEADPHL  240

Qy      240  RVREVCRSHVRIAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFTGLSLTPQL  299
Db      241  RLIREVCRSHVRIAAATLKPPFLRGYRLVRFPEDLAREPLAEIRALYAFTGLTLTPQL  300

Qy      300  EAWHTNTHGSGPGARREAFKTSRNALNVQAWRHALPFAKIRRVQELCAGALQLLGYR  359
Db      301  EAWHTNTHGSGIGKPIAEFTSSRNARNVQAWRHALPFTKILRVQEVCAQALQLLGYR  360

Qy      360  PVYSEDEQRNLALDLVLPRLNGFTWAS  387
Db      361  PVYSAQQRDLTLDLVLPRLNGDHFHSWAS  388

RESULT 3
US-09-949-016-8893
; Sequence 8893, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```


Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTFLLLFLVSRP-----GPSSPAGGEARVHVLVLSWRSRSGSFVQGL 59
 Db 2 LLPKMKLLFLVLSQMAILALFFHMYSHNTSSLSMKQAQPERMHVVLVLSWRSRSGSFVQGL 61
 Qy 60 FNOHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 118
 Db 62 FGQHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 121
 Qy 119 DLFOVAVSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 178
 Db 122 SLFQWNSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 181
 Qy 179 FNLQVLYPLSDPALNLRVHLDPRVAVRSRRTAKALARDNGIVLGTNGTWV-EADP 237
 Db 182 FNLQSLYPLKDPSSLNLHVLVDRPRAVFRSRTKGDLMIDSRVVGQHEQKLKEDQ 241
 Qy 238 GLRVVREVCVSHVRIAEAAATLKP-PPFLGRYRLVRFEDLAREPLAEIRALYFTGLSLT 296
 Db 242 PYYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLLEFL 299
 Qy 297 POLEAWHNTTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRVQELCAGALQLL 356
 Db 300 PHLQVWVHNTTRGKMG--DHAFTNARDALNVSOAWRHSLPYEKVSRLOKACGDAMNLL 357
 Qy 357 GYRPVYSEDEORNLALDLVLPRLGNGFTWASSTASH 392
 Db 358 GYRHVRSEQORNLALLDLL-----STWTVPEQIH 386

RESULT 5

US-09-190-911-1

; Sequence 1, Application US/09190911

; Patent No. 6365365

; GENERAL INFORMATION:

; APPLICANT: Bistrup, Annette

; APPLICANT: Rosen, Steven D.

; APPLICANT: Tangemann, Kirsten

; APPLICANT: Hemmerich, Stefan

; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

; FILE REFERENCE: 6510-107CIP

; CURRENT FILING DATE: 1998-11-12

; EARLIER APPLICATION NUMBER: US/09/190,911

; EARLIER FILING DATE: 1998-03-20

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 1

; LENGTH: 386

; TYPE: PRT

; ORGANISM: H. sapiens

US-09-190-911-1

Query Match 49.6%; Score 1019; DB 3; Length 386;

Best Local Similarity 54.5%; Pred. No. 1.4e-99;

Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

Qy 14 LLLAQTFLLLFLVSRP-----GPSSPAGGEARVHVLVLSWRSRSGSFVQGL 59
 Db 2 LLPKMKLLFLVLSQMAILALFFHMYSHNTSSLSMKQAQPERMHVVLVLSWRSRSGSFVQGL 61
 Qy 60 FNOHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 118
 Db 62 FGQHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 121
 Qy 119 DLFOVAVSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 178
 Db 122 SLFQWNSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 181
 Qy 179 FNLQVLYPLSDPALNLRVHLDPRVAVRSRRTAKALARDNGIVLGTNGTWV-EADP 237
 Db 182 FNLQSLYPLKDPSSLNLHVLVDRPRAVFRSRTKGDLMIDSRVVGQHEQKLKEDQ 241

Qy 238 GLRVVREVCVSHVRIAEAAATLKP-PPFLGRYRLVRFEDLAREPLAEIRALYFTGLSLT 296
 Db 242 PYYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLLEFL 299
 Qy 297 POLEAWHNTTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRVQELCAGALQLL 356
 Db 300 PHLQVWVHNTTRGKMG--DHAFTNARDALNVSOAWRHSLPYEKVSRLOKACGDAMNLL 357
 Qy 357 GYRPVYSEDEORNLALDLVLPRLGNGFTWASSTASH 392
 Db 358 GYRHVRSEQORNLALLDLL-----STWTVPEQIH 386

RESULT 6

US-09-786-240-11

; Sequence 11, Application US/09786240

; Patent No. 6558935

; GENERAL INFORMATION:

; APPLICANT: INCYTE PHARMACEUTICALS, INC.

; APPLICANT: TANG, Y. Tom

; APPLICANT: CORLEY, Neil C.

; APPLICANT: GUEGLER, Karl J.

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: YUE, Henry

; APPLICANT: HILLMAN, Jennifer L.

; APPLICANT: AZIMZAI, Valda

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS

; FILE REFERENCE: PF-0592 PCT

; CURRENT APPLICATION NUMBER: US/09/786,240

; CURRENT FILING DATE: 2002-03-12

; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,642

; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PERL Program

; SEQ ID NO 11

; LENGTH: 386

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1

US-09-786-240-11

Query Match 47.7%; Score 981; DB 4; Length 386;

Best Local Similarity 54.2%; Pred. No. 1.5e-95;

Matches 215; Conservative 42; Mismatches 110; Indels 30; Gaps 9;

Qy 14 LLLAQTFLLLFLVSRP-----GPSSPAGGEARVHVLVLSWRSRSGSFVQGL 59
 Db 2 LLPKMKLLFLVLSQMAILALFFHMYSHNTSSLSMKQAQPERMHVVLVLSWRSRSGSFVQGL 61
 Qy 60 FNOHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 118
 Db 62 FGQHPDVFLYLMPEPAWHVMTTQSQAATLHMVAVDLVRSVFLCDMDVFDAYL-PWRRNLS 121
 Qy 119 DLFOVAVSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 177
 Db 122 SLFQWNSRALCSPACSAFPRGAISSEAVCKPCARQSFTLAREACRSYSHVVLKEVR 180
 Qy 178 FNLQVLYPLSDPALNLRVHLDPRVAVRSRRTAKALARDNGIVLGTNGTWV-EAD 236
 Db 181 FNLQSLYPLKDPSSLNLHVLVDRPRAVFRSRTKGDLMIDSRVVGQHEQKLKEDQ 240
 Qy 237 PGLRVVREVCVSHVRIAEAAATLKP-PPFLGRYRLVRFEDLAREPLAEIRALYFTGLSL 295
 Db 241 PYYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVEFVGLLEFL 298
 Qy 296 TPQLEAWHNTTHGSGPGARREAFKTSRNALNVSOAWRHALPFAKIRVQELCAGALQLL 355
 Db 299 LPHLQVWVHNTTRGKMG--DHAFTNARDALNVSOAWRHSLPYEKVSRLOKACGDAMNLL 356


```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9449

Query Match          32.0%; Score 657.5; DB 4; Length 608;
Best Local Similarity 38.2%; Pred. No. 7.6e-61;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;

Qy 30 GPSPAG-----GEARVHVLLVSSWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTL 80
Db 223 GVAAPCGNTRGTGGVGDKQLVVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTL 282

Qy 81 SQGSAATLHMAVRLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPAC 135
Db 283 YPGDAVSLOGAARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLC 342

Qy 136 SAPPRGAIS--SEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFNQLVLYPLSDPAL 193
Db 343 PAYRKEVVGVLVDVCKK--CPQRLARFEBCRKYRTLVINGVRVFDVAVLAPLRLDPAL 401

Qy 194 NLRIVHLVRDPRAVLRSEOTAKALARDN-----GIVLGT--NGT 231
Db 402 DLKVIHLVRDPRAVASRIRSRHGLIRESLOVVRSDPRAHRMPFLEAGHKLGAKEGV 461

Qy 232 WVEAD--PGLRVVRVCRSHVRIABAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAF 290
Db 462 GGPADYHALGAMEVICNSMAKTLQTA--LQPPDMLOGHVLYVRYEDLVGDPVKTLRRVYDF 520

Qy 291 TGLSLTQLEAHIWNIHSGPGARRBAFKTSSRNALNVSOAHRHALPFAKIRRVQELCA 350
Db 521 VGLLVSPMEQFALNMTSGSSSK--PFVVSARNATQAAWARTALTQQIKQVEEFY 578

Qy 351 GALQLIGYRPVYSDEQNALDLV 375
Db 579 QPMAVLGYERVNSPEEVKDLSTLL 603

RESULT 10
US-09-263-023-4
; Sequence 4, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Rei-ji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-263-023-4

Query Match          31.8%; Score 653.5; DB 3; Length 484;
Best Local Similarity 38.8%; Pred. No. 1.4e-60;
Matches 145; Conservative 70; Mismatches 124; Indels 35; Gaps 10;

```

```

Qy 33 SPAG-GEARVHVLLVSSWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTLSQGSAATLHMA 91
Db 110 APEGVGDKRHHMYVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTLSQGSAATLHMA 169

Qy 92 VRDLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPACSAFPRGAIS-- 144
Db 170 ARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLV 229

Qy 145 SEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDP 204
Db 230 DRAVCKK--CPQRLARFEBCRKYRTLVINGVRVFDVAVLAPLRLDPALDKVIHLVRDP 288

Qy 205 RAVLRSEOTAKALARDN-----GIVLGT--NGTWEAD--PGLRV 241
Db 289 RAVASSRIRSRHGLIRESLOVVRSDPRAHRMPFLEAGHKLGAKEGVGPGADYHALGA 348

Qy 242 RVSVCRSHVRIABAAATLKPPFLRGYRLVRFEDLAREPLAEIRALYAFGLSLTPOLEA 301
Db 349 MEVICNSMAKTLQTA--LQPPDMLOGHVLYVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQ 407

Qy 302 WIHNIHSGPGARRBAFKTSSRNALNVSOAHRHALPFAKIRRVQELCAGALQLLGYRPV 361
Db 408 FALNMTSGSSSK--PFVVSARNATQAAWARTALTQQIKQVEEFYQPMVAVLGYERV 465

Qy 362 YSEDEQNALDLV 375
Db 466 NSPEEVKDLSTLL 479

RESULT 11
US-09-471-867-4
; Sequence 4, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Rei-ji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-471-867-4

Query Match          31.8%; Score 653.5; DB 4; Length 484;
Best Local Similarity 38.8%; Pred. No. 1.4e-60;
Matches 145; Conservative 70; Mismatches 124; Indels 35; Gaps 10;

Qy 33 SPAG-GEARVHVLLVSSWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTLSQGSAATLHMA 91
Db 110 APEGVGDKRHHMYVFTTWSRSGSFVQGLFNQHPDVFYLMPEPAHWVWTTLSQGSAATLHMA 169

Qy 92 VRDLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPACSAFPRGAIS-- 144
Db 170 ARDMLSALYRCDLSVFLQYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGVLV 229

Qy 145 SEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDP 204

```


Qy 314 ARREAFKTSNNALNVSQAWRHAFKIRRVQELCAGALQLLGYRVPVSEQRNIALD 373
 Db 412 S--GIYSTQKNSQSEKFRSMFPKLAQVQAPCGPAMRLFCYKLARDAALNRSVS 468
 Qy 374 LVLPRG 379
 Db 469 LLEERG 474

RESULT 14

US-09-015-188-2
 ; Sequence 2, Application US/09015188C
 ; Patent No. 6393358
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, Kevin J
 ; APPLICANT: Tabas, Ira
 ; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
 ; FILE OF INVENTION: 6-Sulfotransferase
 ; FILE REFERENCE: JEFF-0231
 ; CURRENT APPLICATION NUMBER: US/09/015.188C
 ; CURRENT FILING DATE: 1998-01-29
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 411
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-015-188-2

Query Match 27.0%; Score 555.5; DB 3; Length 411;
 Best Local Similarity 36.7%; Pred. No. 2.9e-50;
 Matches 131; Conservative 56; Mismatches 153; Indels 17; Gaps 8;
 Qy 40 RVHVLSSRRSGSFVQLENQHPDVFYLMPEAWHVVTTL-----ATLHMA 91
 Db 60 KTHILILATTSRSGSFVQLENQHPDVFYLMPEAWHVVTTL-----ATLHMA 119
 Qy 92 VRDLVRSVFLCDMDVFDAYL---PWRRLNLDLQVAVSRALCSPAC--APPRCAISSEA 147
 Db 120 SRDLRSVFLCDMDVFDAYL---PWRRLNLDLQVAVSRALCSPAC--APPRCAISSEA 179
 Qy 148 VCKPLCARQSTFLAREACRSYSHVVLKEVPFNFQVLYPLSLDLPALNLRVHVLVDRPRAV 207
 Db 180 DCVRKCGLLNLTVAAEACRSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVDRPRAV 239
 Qy 208 LRSREQTAKALARDNGIVLGTNGTWREADPGLRVVREVCRRSHVRIAEATLKPFPFLGR 267
 Db 240 LASRSETPRDYRLWRLWYGTGRKPYNLD--VTQLTTVCEDFSNSVSTGLMR--PPWLKKG 296
 Qy 268 YRLVRFEDLAREPLAEIRALYAFGLSLTPOLEAWIHNITHGSGGARREAFKTSNNAL 327
 Db 297 YMLVRYEDLARNPKMKTTEEYVGLIGIPLDSHVARIQNTTGRD--PTLGHKHYGT--VRNSA 354
 Qy 328 NVSQAWRHAFKIRRVQELCAGALQLLGYRVPVSEQRNIALDLVLPRGLNGFT 384
 Db 355 ATAEXRFLSYDVAFAQNAQQVLAQGLGYKIAASEELKNPSVLSVEERDFRPS 411

RESULT 15

US-08-655-878-2
 ; Sequence 2, Application US/08655878
 ; Patent No. 582713
 ; GENERAL INFORMATION:
 ; APPLICANT: FUKUTA, MASAKAZU
 ; APPLICANT: HABUCHI, OSAMI
 ; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE:
 ; STREET:
 ; CITY:
 ; STATE:
 ; COUNTRY:

; ZIP:
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/655,878
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME:
 ; REGISTRATION NUMBER:
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE:
 ; TELEFAX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 458
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-655-878-2

Query Match 25.9%; Score 532.5; DB 2; Length 458;
 Best Local Similarity 34.6%; Pred. No. 9.7e-48;
 Matches 134; Conservative 63; Mismatches 157; Indels 33; Gaps 10;
 Qy 5 RVSVTAVTALLAQTFLLFLVSRPGPSSPAGGEARVHVLVLSWRSRSGSFVQLENQHP 64
 Db 80 RQDTHAFSQRRLRNVTQLAGELGIAAP---EPRRHVLLMATTGSSSFVGEFFNQOQ 136
 Qy 65 DVFLMEPAWHV---WTTLSQSSAATLHMAV---RDLVRSVFLCDMDVFDAYL---PWRRLN 117
 Db 137 NIFLFEPLWHIERTVTFEPGGANAVGSALVYRDVLQQLLCLDLYLEISFISPAPEH 196
 Qy 118 SDFQWAVSRALCSPACSAFPRAISSEAVCKPLCARQSTFLAREACRSYSHVVLKEVR 177
 Db 197 AALFRGSSHSLCEEPVCTPSLKKVFEKYHCKNRCGLNITLAAEACRRKQHWALKTVR 256
 Qy 178 FFNLQVLYPLSLDLPALNLRVHVLVSRPRAVLRSEQTAKALARDNGIVLGTNGTW---V 233
 Db 257 IROLEFLQPLAEDPRLDLRIQLVDRPRAVLRVSRWAFS-----GKYESKWKAA 306
 Qy 234 EADPGLRV-----VREVCRRSHVRIAEATLKPFPFLGRYRLVRFEDLAREPLAEIRALY 288
 Db 307 EGEAPLQEDDEVQRLGNCS-IRLSAELGLRQPRWLRGRLVRYEDVARAPLRKALEMY 365
 Qy 289 AFTGLSLTPOLEAWIHNITHGSGGARREAFKTSNNALNVSQAWRHAFKIRRVQEL 348
 Db 366 RFAGIHPTQVEEWIRANT--QAQDSNGIYST--QKNSSEQFEKWRFSIPPKLAQVQVDA 422
 Qy 349 CAGALQLLGYRVPVSEQRNIALDLV 375
 Db 423 CEPAMRLFGYKLASSAQELNRLSL 449

Search completed: June 23, 2005, 08:52:14
 Job time : 15.8883 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 49.935 Seconds
(without alignments)
3041.886 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10A_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2056	100.0	395	9	US-09-927-602-2
2	2056	100.0	395	15	US-10-258-080-5
3	2056	100.0	395	16	US-10-648-593-159
4	2056	100.0	395	16	US-10-697-828-13
5	1934	94.1	395	9	US-09-927-602-3
6	1729.5	84.1	390	9	US-09-927-602-4
7	1729.5	84.1	390	16	US-10-697-828-8
8	1708	83.1	395	16	US-10-697-828-7
9	1708	83.1	395	16	US-10-841-707-8
10	1694.5	82.4	418	9	US-09-927-602-5
11	1538	74.8	394	16	US-10-408-765A-2211

12	1019	49.6	386	9	US-09-916-825-2	Sequence 2, Appli
13	1019	49.6	386	13	US-10-007-262-1	Sequence 1, Appli
14	1018	49.5	380	16	US-10-841-707-6	Sequence 6, Appli
15	981	47.7	386	14	US-10-427-631-11	Sequence 11, Appli
16	808	39.3	169	9	US-09-927-602-6	Sequence 6, Appli
17	755	36.7	169	9	US-09-927-602-7	Sequence 7, Appli
18	663	32.2	483	14	US-10-212-933-2	Sequence 2, Appli
19	657.5	32.0	530	16	US-10-723-860-1409	Sequence 1409, Ap
20	653.5	31.8	484	14	US-10-212-933-4	Sequence 4, Appli
21	653.5	31.8	531	9	US-09-933-790-255	Sequence 255, App
22	653.5	31.8	531	16	US-10-755-889-284	Sequence 284, App
23	578.5	28.1	479	13	US-10-087-192-126	Sequence 126, App
24	569	27.7	481	13	US-10-087-192-123	Sequence 123, App
25	555.5	27.0	411	14	US-10-021-660-128	Sequence 128, App
26	555.5	27.0	411	15	US-10-211-462-97	Sequence 97, Appli
27	555.5	27.0	411	16	US-10-408-765A-395	Sequence 395, App
28	555.5	27.0	411	16	US-10-723-860-1544	Sequence 1544, Ap
29	521.5	25.4	171	9	US-09-927-602-8	Sequence 8, Appli
30	377	18.3	169	9	US-09-927-602-9	Sequence 9, Appli
31	364.5	17.7	174	9	US-09-927-602-11	Sequence 11, Appli
32	348	16.9	179	9	US-09-927-602-10	Sequence 10, Appli
33	328	16.2	387	14	US-10-126-279-21	Sequence 21, Appli
34	328	16.2	387	14	US-10-286-606-21	Sequence 21, Appli
35	328	16.2	387	16	US-10-891-383-21	Sequence 21, Appli
36	123.5	6.0	596	16	US-10-697-828-9	Sequence 9, Appli
37	123.5	6.0	1212	17	US-10-479-472A-2	Sequence 2, Appli
38	123.5	6.0	1222	16	US-10-697-828-15	Sequence 15, Appli
39	123.5	6.0	1223	16	US-10-475-446-4	Sequence 4, Appli
40	118.5	5.8	1207	16	US-10-697-828-17	Sequence 17, Appli
41	118.5	5.8	1207	17	US-10-479-472A-4	Sequence 4, Appli
42	108	5.3	2588	16	US-10-437-963-138279	Sequence 138279, A
43	107	5.2	299	14	US-10-156-761-9846	Sequence 9846, Ap
44	98	4.8	276	15	US-10-425-114-47175	Sequence 47175, A
45	98	4.8	281	16	US-10-425-115-213557	Sequence 213557, A

ALIGNMENTS

RESULT 1
US-09-927-602-2
; Sequence 2, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; FILE REFERENCE: P-LJ 4852
; CURRENT FILING DATE: 2001-08-09
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-2

Query Match	100.0%	Score 2056;	DB 9;	Length 395;
Best Local Similarity	100.0%	Pred. No. 4.6e-198;		
Matches 395;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MWLPVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSQSGSFVQGLF	60	
Db	1	MWLPVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSQSGSFVQGLF	60	
Qy	61	NQHPDVFYLMPEAWHVVTTLSQSSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL	120	
Db	61	NQHPDVFYLMPEAWHVVTTLSQSSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL	120	

Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
 Qy 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Db 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Qy 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Db 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
 Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 4

US-10-697-828-13
 ; Sequence 13, Application US/10697828
 ; Publication No. US20040185546A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, Steven
 ; APPLICANT: Lee, Jin Kyu
 ; APPLICANT: Hemmerich, Stefan
 ; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
 ; FILE REFERENCE: UCAL-138DIV
 ; CURRENT APPLICATION NUMBER: US/10/697,828
 ; CURRENT FILING DATE: 2003-10-29
 ; PRIOR APPLICATION NUMBER: 09/593,828
 ; PRIOR FILING DATE: 2000-06-13
 ; PRIOR APPLICATION NUMBER: 60/144,694
 ; PRIOR FILING DATE: 1999-07-20
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 13
 ; LENGTH: 395
 ; TYPE: PRT
 ; ORGANISM: homo sapiens
 US-10-697-828-13

Query Match 100.0%; Score 2056; DB 16; Length 395;
 Best Local Similarity 100.0%; Pred. No. 4.6e-198;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGVGLF 60
 Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGVGLF 60
 Qy 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
 Db 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
 Qy 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 Db 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 Qy 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
 Qy 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Db 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Qy 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Db 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
 Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395

RESULT 5

US-09-927-602-3
 ; Sequence 3, Application US/09927602
 ; Patent No. US20020061562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; APPLICANT: Akama, Tomoya O.
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal
 ; FILE REFERENCE: P-LJ 4852
 ; CURRENT APPLICATION NUMBER: US/09/927,602
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: US 09/638,211
 ; PRIOR FILING DATE: 2000-08-11
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 395
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: synthetic construct
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(395)
 ; OTHER INFORMATION: Xaa = any amino acid
 US-09-927-602-3

Query Match 94.1%; Score 1934; DB 9; Length 395;
 Best Local Similarity 94.4%; Pred. No. 9.1e-186;
 Matches 371; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
 Qy 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGVGLF 60
 Db 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVLVLSWRSWSSGSSFGVGLF 60
 Qy 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
 Db 61 NQHPDVFYLMPEPAWHVMTTLLSQGSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
 Qy 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 Db 121 FQWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRFN 180
 Qy 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
 Db 181 LQVLYPLSDPALNLRIVHLVRDPAVLRSEQTAKALARDNGIVLGTNGTWVADPGLR 240
 Qy 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Db 241 VVREVCVSHVRIAEAAATLKPPFLGRVRLVRFEDLAREPLAEIRALYAFGLSLTPGLE 300
 Qy 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Db 301 AWIHNITHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLGYRP 360
 Qy 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 393
 Db 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 393

RESULT 6

US-09-927-602-4
 ; Sequence 4, Application US/09927602
 ; Patent No. US20020061562A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; APPLICANT: Akama, Tomoya O.
 ; TITLE OF INVENTION: Methods of Treating Macular Corneal
 ; FILE REFERENCE: P-LJ 4852
 ; CURRENT APPLICATION NUMBER: US/09/927,602
 ; CURRENT FILING DATE: 2001-08-09
 ; PRIOR APPLICATION NUMBER: US 09/638,211
 ; PRIOR FILING DATE: 2000-08-11

; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 84.1%; Score 1729.5; DB 9; Length 390;
Best Local Similarity 85.8%; Pred. No. 3.6e-165;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLVSSWRSGSSFVQGL 59
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLVSSWRSGSSFVQGL 60
Qy 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYLWVRNLS 119
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYMPOSRNLSA 120
Qy 120 LFOAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRPF 179
Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCRTQPPSLAREACRSYSHVVLKEVRPF 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEQAAGPILARDNGIVLGTNGKWVEADPHL 240
Qy 240 RVREVCVSHVRIAEATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLSLTPQL 299
Db 241 RLIREVCVSHVRIAEATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLSLTPQL 300
Qy 300 EAWIHNTHGSGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 359
Db 301 EAWIHNTHGSGIGKPIEAFTTSRNARNVSOAQRHALPFTKILRVQEVCGALQLLYR 360
Qy 360 PVYSEDEQORNALDLVLRGLNGFTWAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 7

US-10-697-828-8
; Sequence 8, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 390
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-8

Query Match 84.1%; Score 1729.5; DB 16; Length 390;
Best Local Similarity 85.8%; Pred. No. 3.6e-165;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

Qy 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLVSSWRSGSSFVQGL 59
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLVSSWRSGSSFVQGL 60

Qy 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYLWVRNLS 119
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYMPOSRNLSA 120
Qy 120 LFOAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRPF 179
Db 121 FFWNATSRALCSPACSAFPRGTISKQDVCKTLCRTQPPSLAREACRSYSHVVLKEVRPF 180
Qy 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 239
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSEQAAGPILARDNGIVLGTNGKWVEADPHL 240
Qy 240 RVREVCVSHVRIAEATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLSLTPQL 299
Db 241 RLIREVCVSHVRIAEATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLSLTPQL 300
Qy 300 EAWIHNTHGSGPGARREAFKTSRNALNVSOAQRHALPFAKIRRVQELCAGALQLLYR 359
Db 301 EAWIHNTHGSGIGKPIEAFTTSRNARNVSOAQRHALPFTKILRVQEVCGALQLLYR 360
Qy 360 PVYSEDEQORNALDLVLRGLNGFTWAS 387
Db 361 PVYSADQQRDLTLVLPRGPDHFSWAS 388

RESULT 8

US-10-697-828-7
; Sequence 7, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 395
; TYPE: PRT
; ORGANISM: mus musculus
US-10-697-828-7

Query Match 83.1%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.4e-163;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MWLPRVSTAVTALLAQ-TFLLLFLVSRPGSPAGGEARVHVLVSSWRSGSSFVQGL 60
Db 1 MWLPRFSKTVTVLLLAQTCLLLFIISRPQSPAGGEDRVHVLVSSWRSGSSFVQGL 60
Qy 61 FNOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYLWVRNLS 120
Db 61 FSOHPDVFLMEPAWHVWTTLSQGSAAATHMAVRDLVRSVFLCDMDVDFDAYLWVRNLS 120
Qy 121 FOWAVSRALCSPACSAFPRGAISSSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRPF 180
Db 121 FOWAVSRALCSPVCEAFARGNISSEEVCKPLCATRPFGLAQECSSYSHVVLKEVRPF 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRAVLRSEQTAKALARDNGIVLGTNGTWVEADPGL 240
Qy 241 VVREVCVSHVRIAEATLKPFPFLRGYRLVRFEDLAREPLAEIRALYAFTGLSLTPQL 300
Db 241 VVNEVCVSHVRIAEATLKPFPFLQDRYLRVEDLARDPLTVIRELYAFTGLSLTPQL 300

```

Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLLYR 360
Db 301 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYR 360
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VHSLEQDLSDLLPRGMDSFKWASSTEKOPES 395

RESULT 9
US-10-841-707-8
; Sequence 8, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru
; APPLICANT: Yeh, Jiunn-Chern
; APPLICANT: Hiraoka, Nobuyoshi
; TITLE OF INVENTION: Identification of the Meca-79 Antigen
; TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
; TITLE OF INVENTION: Conditions
; FILE REFERENCE: P-LJ 4149
; CURRENT APPLICATION NUMBER: US/10/841,707
; CURRENT FILING DATE: 2004-05-05
; PRIOR APPLICATION NUMBER: US/09/569,320A
; PRIOR FILING DATE: 2000-05-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-841-707-8

Query Match 83.1%; Score 1708; DB 16; Length 395;
Best Local Similarity 83.0%; Pred. No. 5.4e-163;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

Qy 1 MWLPRVSVTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSWGSSFGVGLF 60
Db 1 MRLPRFSSTVMSLLMVQTGILVLSRQVPSPAGLGERVHVLVLSWRSWGSSFGVGLF 60
Qy 61 NQHPDVFLMEPAWVHTTSLQGSAAHLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 61 SQHPDVFLMEPAWVHTTSLQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNLSDL 120
Qy 121 FQWAVSRALCSPPCSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
Db 121 FQWAVSRALCSPPCSAFPFGAISSEAVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 180
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
Db 181 LQVLYPLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
Qy 241 VVREVCRSVHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Db 241 VVREVCRSVHRIAEEATLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFGLSLTPGLE 300
Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLLYR 360
Db 301 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYR 360
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 361 VHSLEQDLSDLLPRGMDSFKWASSTEKOPES 395

RESULT 10
US-09-927-602-5
; Sequence 5, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.

```

```

; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-927-602-5

Query Match 82.4%; Score 1694.5; DB 9; Length 418;
Best Local Similarity 82.8%; Pred. No. 1.3e-161;
Matches 327; Conservative 24; Mismatches 43; Indels 1; Gaps 1;

Qy 1 MWLPRVSVTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWRSWGSSFGVGLF 60
Db 25 MRLPRFSSTVMSLLMVQTGILVLSRQVPSPAGLGERVHVLVLSWRSWGSSFGVGLF 84
Qy 61 NQHPDVFLMEPAWVHTTSLQGSAAHLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDL 120
Db 85 SQHPDVFLMEPAWVHTTSLQGSAPALHMAVRDLIRSVFLCDMDVFDAYLPWRNLSDL 144
Qy 121 FQWAVSRALCSPPCSAFPFGAISSEAVCKPLCARQSFLLAREACRSYSHVVLKEVRF 180
Db 145 FQWAVSRALCSPPCSAFPFGAISSEAVCKPLCATRPFGLAQEACSSYSHVVLKEVRF 204
Qy 181 LQVLYPLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
Db 205 LQVLYPLSDPALNLRIVHLVRDPRAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 264
Qy 241 VVREVCRSVHRIAEEATLKPPPLRGYRLVRFPEDLAREPLAEIRALYAFGLSLTPGLE 300
Db 265 VVREVCRSVHRIAEEATLKPPPLQDRYRLVRYEDLARDPLTVIRELYAFGLSLTPGLE 323
Qy 301 AWIHNTHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQELCAGALQLLYR 360
Db 324 TWIHNTHGSGPGARREAFKTSRDALSVQAWRHLPFAKIRRVQELCAGALQLLYR 383
Qy 361 VYSEDEQNLALDLVLPRLNGFTWASSTASHPRN 395
Db 384 VHSLEQDLSDLLPRGMDSFKWASSTEKOPES 418

RESULT 11
US-10-408-765A-2211
; Sequence 2211, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2211
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2211

```

Query Match	74.8%;	Score 1538;	DB 16;	Length 394;
Best Local Similarity	93.3%;	Pred. No. 7.2e-146;		
Matches 294;	Conservative 7;	Mismatches 14;	Indels 0;	Gaps 0;
Qy 1	MMLPRVSSTAVTALLLAQTLLFLVSRPGPSSPAGGEARVHVILVLSWSRSGSFVGOLF	60		
Db 1	MMLPRVSSTAVTALLLAQTLLFLVSRPGPSSPAGGEARVHVILVLSWSRSGSFVGOLF	60		
Qy 61	NQHPDVFYLMPEAWHVMTTSLSQGSAATLHMVARDLVRVSFLCDMDVFDAYLPWRNLSDL	120		
Db 61	NQHPDVFYLMPEAWHVMTTSLSQGSAATLHMVARDLVRVSFLCDMDVFDAYLPWRNLSDL	120		
Qy 121	FQWAVSPALCSPPACSAFPRGAISSSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFN	180		
Db 121	FQWAVSPALCSPPACSAFPRGAISSQDVCKTLCARQFPFSLAREACRSYSHVVLKEVRFN	180		
Qy 181	LOVLYPLLSPALNLRIVHLVRDPRAVLRSFEQAKALARNGIVLGTNGTWVADPCLR	240		
Db 181	LOVLYPLLSPALNLRIVHLVRDPRAVLRSFEQAKALARNGIVLGTNGTWVADPCLR	240		
Qy 241	VIREVCKSHVRIAEAAATLKPPPFILRGYRLVRFFEDLAREPLAEIRALYAFYFTGLSILTPOLE	300		
Db 241	LIREVCESHVRIAEAAATLKPPPFILRGYRLVRFFEDLAREPLAEIRALYAFYFTGLTITPQLE	300		
Qy 301	AWIHNTHGSGPGAR	315		
Db 301	AWIHNTHGRGSASQ	315		

```

RESULT 12
US-09-816-825-2
; Sequence 2, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-825-2

```

```

Db      242  PYVMQVICSQLEIYK--TIQSLPKALQERYLLVRVEDIARAPVAQTSRMVEFVGLEFL  299
Qy      297  POLEAWTHNITHGSGPGARREAFKTSRNALNVSAQWRHALPFKIRRVQELCAGALQLL  356
Db      300  PHLQTVHVNITRGKMG- -DHAFHTNARDALNVSAQWRSLPYEKVSR/LQACGDAMNLL  357
Qy      357  GYRPVYSEQRNLALDLVLPRGLNGETWASSTASH  392
Db      358  GYRHRVSEQRNLLDLL-----STWTVPEQIH  386

RESULT 13
US-10-007-262-1
; Sequence 1, Application US/10007262
; Publication NO. US20020184748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-10-007-262-1

```

RESULT 14
US-10-841-707-6
; Sequence 6, Application US/10841707
; Publication No. US20040202649A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Minoru

APPLICANT: Yeh, Jiunn-Chern
APPLICANT: Hiraoka, Nobuyoshi
TITLE OF INVENTION: Identification of the Meca-79 Antigen
TITLE OF INVENTION: and Related Methods of Treating L-Selectin-Mediated
TITLE OF INVENTION: Conditions
FILE REFERENCE: P-LJ 4149
CURRENT APPLICATION NUMBER: US/10/841,707
CURRENT FILING DATE: 2004-05-06
PRIOR APPLICATION NUMBER: US/09/569,320A
FILING FILING DATE: 2000-05-11
NUMBER OF SEQ ID NOS: 29
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 6
LENGTH: 380
TYPE: PRT
ORGANISM: Homo Sapien
US-10-841-707-6

Query Match 49.5%; Score 1018; DB 16; Length 380;
Best Local Similarity 55.0%; Pred. No. 1.7e-93;
Matches 214; Conservative 44; Mismatches 103; Indels 28; Gaps 7;

Qy 21 LLLFLVSRP-----CPSSPAGGEARVHVLVSSWSSGSSFGQLFQHPDV 66
Db 3 LLLFLVSRP-----CPSSPAGGEARVHVLVSSWSSGSSFGQLFQHPDV 62

Qy 67 FYLMEPAHWVMTTSLQSSAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLSDLFQWAV 125
Db 63 FYLMEPAHWVMTTSLQSSAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLSDLFQWAV 122

Qy 126 SRALCSPPACSAFPRGAISSBAVCKPLCARQSFLLAREACRSYSHVVLKEVRFNLOVLY 185
Db 123 SRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVEKACRSYSHVVLKEVRFNLOVLY 182

Qy 186 PLLSDPALNLRIVLVRDPRVLRSEOTAKALARDNGIVLTNGTWV-BADPGLRVVRE 244
Db 183 PLLSDPALNLRIVLVRDPRVLRSEOTAKALARDNGIVLTNGTWV-BADPGLRVVRE 242

Qy 245 VCRSHVRIAEAAATLKP-PPLRGRVRLVRFEDLAREPLAEIRALYAFGLSLTPQLEAWI 303
Db 243 ICQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLFLEPHLQTWV 300

Qy 304 HNITHGSGPGARREAFKTSRRNALNVSQAWRHLPFAKIRRVQELCAGALQLLGYRPPVYS 363
Db 301 HNITHGSGPGARREAFKTSRRNALNVSQAWRHLPFAKIRRVQELCAGALQLLGYRPPVYS 358

Qy 364 EDEORNALDLVLRGLNGFTWASSTASH 392
Db 359 EDEORNALDLVLRGLNGFTWASSTASH 380

RESULT 15
US-10-427-631-11
Sequence 11, Application US/10427631
Publication No. US20030175923A1
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; TANG, Y. Tom;
APPLICANT: CORLEY, Neil C.; GUEGLER, Karl J.;
APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
APPLICANT: YUE, Henry; HILLMAN, Jennifer L.;
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PF-0592-1 DIV
CURRENT APPLICATION NUMBER: US/10/427,631
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 09/786,240
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: PCT/US99/20989
PRIOR FILING DATE: 1999-09-09
PRIOR APPLICATION NUMBER: US 60/172,220
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: US 60/155,248
PRIOR FILING DATE: 1998-11-04

PRIOR APPLICATION NUMBER: US 60/133,642
PRIOR FILING DATE: 1999-05-11
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PERL Program
SEQ ID NO 11
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030175923A1 2617407CD1
US-10-427-631-11

Query Match 47.7%; Score 981; DB 14; Length 386;
Best Local Similarity 54.2%; Pred. No. 9.4e-90;
Matches 215; Conservative 42; Mismatches 110; Indels 30; Gaps 9;

Qy 14 LLLAQTFLLFLVSRP-----GPSSPAGGEARVHVLVSSWSSGSSFGQL 59
Db 2 LLLPKMKLLFLVSRP-----GPSSPAGGEARVHVLVSSWSSGSSFGQL 61

Qy 60 FNOHPDVFYLMPEPAHWVMTTSLQSSAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 118
Db 62 FGOHPDVFYLMPEPAHWVMTTSLQSSAATLHMAVRDLVRSVFLCDMDVFDAYL-PWRRNLS 121

Qy 119 DLFQWAVSRALCSPPACSAFPRGAISSBAVCKPLCARQSFLLAREACRSYSHVVLKEVR 177
Db 122 SLFQWENSRALCSAPACDIIPQDEIIPRAHCRLLCSQPFVEVEKACRSYSHVVLKEVR 180

Qy 178 FFLQVLYPLSDPALNLRIVLVRDPRVLRSEOTAKALARDNGIVLTNGTWV-BAD 236
Db 181 FFLQVLYPLSDPALNLRIVLVRDPRVLRSEOTAKALARDNGIVLTNGTWV-BAD 240

Qy 237 PGLRVVREVCRSYSHVRIAEAAATLKP-PPLRGRVRLVRFEDLAREPLAEIRALYAFGLSL 295
Db 241 QPYVMQVICSQSOLEIYK--TIQSLPKALQERYLLVRYEDLARAPVAQTSRMVFEVGLF 298

Qy 296 TPQLEAWIHNIHITHGSGPGARREAFKTSRRNALNVSQAWRHLPFAKIRRVQELCAGALQL 355
Db 299 LPHLQTWVHNITRGKMG--DHAFHTNARDALNVSQAWRHLPFAKIRRVQELCAGALQL 356

Qy 356 LGYRPPVYSEDEORNALDLVLRGLNGFTWASSTASH 392
Db 357 LGYRPPVYSEDEORNALDLVLRGLNGFTWASSTASH 386

Search completed: June 23, 2005, 09:27:57
Job time : 51.935 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 8.95931 Seconds
(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-13
Perfect score: 2056
Sequence: 1 MMLPRVSSSTAVTALLLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	656.5	31.9	484	2 JE0261	N-acetylglucosamin
2	618.5	30.1	484	2 JC7350	N-acetylglucosamin
3	615	29.9	486	2 JC7351	N-acetylglucosamin
4	532.5	25.9	458	2 A57397	chondroitin 6-sulf
5	128	6.2	388	2 G70729	hypothetical prote
6	118	5.7	307	2 E95934	probable enzyme, C
7	100.5	4.9	523	2 B83629	probable ATP-bindi
8	98	4.8	559	2 T03412	malate synthase (E
9	93	4.5	359	2 T16350	hypothetical prote
10	91.5	4.5	425	2 E83023	3-deoxy-D-manno-oc
11	91.5	4.5	577	2 B75595	probable long-chain
12	91.5	4.5	617	2 F82744	ferrous iron trans
13	91.5	4.5	820	2 H86246	hypothetical prote
14	91	4.4	380	2 A75328	phoH-related prote
15	90.5	4.4	349	2 E75611	glucosamine-fructo
16	90.5	4.4	710	1 A46273	mitochondrial inte
17	90	4.4	931	2 AF3276	ATP-dependent clp
18	89.5	4.4	221	2 T50665	hypothetical prote
19	89.5	4.4	668	2 T01695	clp1 protein - mai
20	89.5	4.4	1489	2 S73015	polyketide synthas
21	89.5	4.4	3643	2 T36410	probable polyketid
22	89	4.3	571	2 AG3174	acetolactate synth
23	88.5	4.3	262	2 G95327	Trm3 transposase f
24	88.5	4.3	400	2 S24759	probable transposa
25	88.5	4.3	400	2 A42727	probable transposa
26	88.5	4.3	400	2 C95306	Trm3 transposase [
27	88.5	4.3	400	2 E95287	Trm3 transposase [
28	88.5	4.3	400	2 H95291	Trm3 transposase [
29	88.5	4.3	400	2 P95354	Trm3 transposase [

30 88.5 4.3 809 2 C71279 probable ATP-depen
31 88 4.3 370 2 A49360 pyruvate dehydroge
32 88 4.3 390 2 G82844 cysteine synthase
33 87.5 4.3 260 2 B95965 probable two-compo
34 87.5 4.3 467 2 E95981 conserved hypothet
35 87.5 4.3 547 1 S51475 cytochrome P450 Cy
36 87 4.2 327 2 T50744 spheroidene monoox
37 87 4.2 389 1 DEPGA pyruvate dehydroge
38 87 4.2 532 2 S33758 flavin-containing
39 86.5 4.2 570 2 G95326 probable acetolact
40 86.5 4.2 856 2 F87316 conserved hypothet
41 86 4.2 528 2 T51119 phytoene desaturas
42 86 4.2 652 2 T45499 probable DNA recom
43 86 4.2 787 2 AB3207 conserved hypothet
44 85 4.1 293 2 C71267 probable carboxyle
45 85 4.1 327 2 S49619 crtA protein - Rho

ALIGNMENTS

RESULT 1
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: JE0261
R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.
J. Biochem. 124, 670-678, 1998
A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of f
A:Reference number: JE0261; MUID:98391845; PMID:9722682
A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenos
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match	31.9%	Score	656.5	DB 2	Length	484			
Best Local Similarity	39.0%	Pred. No.	5.9e-50						
Matches	146	Conservative	69	Mismatches	124	Indels	35	Gaps	10
Qy	33	SPAG-GEARVHVVLVLSWSRSGSVFQGFNQHPDVFYLMEPAWHVMTTSLQSGSAATHMA	91						
Db	110	APEGVGDKRHWYVFTTWKSGSGSFFGELFNQNPVEFFLYEPVHWVWQKLYPGDAVSQGA	169						
Qy	92	VRDLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPACSAFPRGALS--	144						
Db	170	ARDMLSALYRCDLSVFLQYSPAGSGGNLTTLGIFGAATNKVWCSSPLCPAYRKEVVGLV	229						
Qy	145	SEAVCKPLCARQSFTHLAREACRSYHVLVKEVFFNLOVLYPLSLDPAALNLRIVHLVRDP	204						
Db	230	DDRVCCK-CPQRLARPEECCRYRTLKGVRFVDAVLAPLIRDPALDLKVTHLVRDP	288						
Qy	205	RAVLSRBEQTAKALARDN-----GIVLGT--NGTWVEAD-PGLRV	241						
Db	289	RAVASSRIRSRHGLIRSLQVRSRDPRAHRMPFLEAAGHKJGAKKEGVGGPDYHALGA	348						
Qy	242	VREVCVSHVRIAEATLKPPFLRGVRLRVRFEDLAREPLAEIRALYAFGLSLTPOLEA	301						
Db	349	MEVICHSMATLQTA-LQPPDWLQGHVLRVEDLVCEPVKTLRRVYDFVGLLVSPMEQ	407						
Qy	302	WIHNTHGSGPGARAEAFKTSRNALNVSOAQRHALPFKAKIRRVQELCAGALQLLGRPV	361						
Db	408	FALNMTSGSGSSK--PFVVSARNATQAAANWRTALTFOQIKQVEEFCYOPMAVLGVERV	465						
Qy	362	YSEDEQRNALDLV	375						
Db	466	NSPEEVKDLSTLL	479						

RESULT 2
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C;Accession: JC7350
R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7350
A;Molecule type: mRNA
A;Residues: 1-484 <UCH>
A;Cross-references: UNIPROT:Q99NB0; UNIPROT:Q9EP78; DDBJ:AB040710
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.1%; Score 618.5; DB 2; Length 484;
Best Local Similarity 38.0%; Pred. No. 1.3e-46;
Matches 148; Conservative 56; Mismatches 130; Indels 55; Gaps 11;

QY 31 PSSP-----ACGEA----RVHVLVSSWRSGSFFVGOLFNOHPDVFYLMPEPAHWVTTLSQ 82
DB 82 PGSPGNLSAAGEAVTOEKQHYVHATWRTGSSFLGELFNOHPDVFYLYEPMHLMQALYLP 141
QY 83 GSAATLHMVDRVRSVFLCDMDVFDAY-----LPMRNLSD--LFWAVSRALCS 131
DB 142 GDAESLQALRDLMLRSIFRCDFSVLRLYAQPDGGERAPDSANLTTAMLFRWRTNKVICS 201
QY 132 PPACSAFPR-----GAISSEAVCKPFCARQSFFLAREACRSYSHVVLKEVRFENLOVLYP 186
DB 202 PPLCPAAPARADVGLVEDKA-CESTCPPVLSRLAEAECKRYPVVVIKDVRLDLGLVPLR 260
QY 187 LLSDPALNLRIHVLRDPRAVLRSEQTAKALARDNGIVLGTN-----GTWVEADPG 235
DB 261 LLADPGLNLKVQLFRDPRAVHNSRLKSRQLRESIQVLTQRGDHFHRLVLAHGVD 320
QY 236 DPG-----LRVREVCRSYHVRVIAEATLKPPFPLRGYRLVRFPEDLARE 279
DB 321 RPPGQARALPSAPRADFFLTSALEVICENWLRDLJLFARGA----PAWLRRYRLRYEDLV 379
QY 280 PLAEIRALYFTGLSLTPQLAEAWHITHGSGPGARREAFKTSRRNALNVSOAHRHALPF 339
DB 380 PQQLRRLRFSGRLTALDAFALNMTGAAYGADR-PFHLSARDAREAVHWRERLSR 438
QY 340 AKIRRVQELCAGALQLLGY 368
DB 439 EQVRQVEACAPAMRLLAY-PRSGDERDR 466

RESULT 3
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C;Species: Homo sapiens (man)
C;Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C;Accession: JC7351
R;Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A;Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A;Reference number: JC7350; MUID:20374462; PMID:10913333
A;Accession: JC7351
A;Molecule type: mRNA
A;Residues: 1-486 <UCH>
A;Cross-references: UNIPROT:Q75667; DDBJ:AB040711
C;Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C;Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.9%; Score 615; DB 2; Length 486;
Best Local Similarity 38.0%; Pred. No. 2.7e-46;
Matches 144; Conservative 53; Mismatches 128; Indels 54; Gaps 8;

QY 29 PGSPSAGGEA----RVHVLVSSWRSGSFFVGOLFNOHPDVFYLMPEPAHWVTTLSQGS 84
DB 86 PSNLSGAVGEAVSRKQHYVHATWRTGSSFLGELFNOHPDVFYLYEPMHLMQALYPGD 145
QY 85 AATLHMVDRVRSVFLCDMDVFDAY-----LPMRNLSD--LFWAVSRALCSPP 133
DB 146 AESLQALRDLMLRSIFRCDFSVLRLYAPPGDPAARADPTANLTTAALFRWRTNKVICSPP 205
QY 134 ACSAPPRG-----AISSEAVCKPFCARQSFFLAREACRSYSHVVLKEVRFENLOVLYPLS 189
DB 206 LCPGAPRAREAVGLVEDTACERSCPPVAIRALEAECKRYPVVVIKDVRLDLGLVPLLR 265
QY 190 DPALNLRIHVLRDPRAVLRSEQTAKALARDNGIVLGTN-----GTWVEADPG 238
DB 266 DPGLNLKVQLFRDPRAVHNSRLKSRQLRESIQVLTQRGDHFHRLVLAHGVGARPG 325
QY 239 -----LRVREVCRSYHVRVIAEATLKPPFPLRGYRLVRFPEDLARE 279
DB 326 QOSRALPAAPRADFFLTGALVICENWLRDLJLFARGA----PAWLRRYRLRYEDLV 381
QY 280 PLAEIRALYFTGLSLTPQLAEAWHITHGSGPGARREAFKTSRRNALNVSOAHRHALPF 339
DB 382 PRAQLRRLRFSGRLTALDAFALNMTGAAYGADR-PFHLSARDAREAVHWRERLSR 440
QY 340 AKIRRVQELCAGALQLLGY 358
DB 441 EQVRQVEACAPAMRLLAY 459

RESULT 4
A57397
Chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004
C;Accession: A57397
J;Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Habuchi,
J. Biol. Chem. 270, 18575-18580, 1995
A;Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransf
A;Reference number: A57397; MUID:95355490; PMID:7629189
A;Accession: A57397
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-458 <FUK>
A;Cross-references: UNIPROT:Q92179; GB:D49915; NID:g971262; PIDN:BAA08655.1; PID:g971263
C;Superfamily: chondroitin 6-sulfotransferase
C;Keywords: sulfotransferase

Query Match 25.9%; Score 532.5; DB 2; Length 458;
Best Local Similarity 34.6%; Pred. No. 4.9e-39;
Matches 134; Conservative 63; Mismatches 157; Indels 33; Gaps 10;

QY 5 RVSSTAVTALLAQTFLLFLVSRPGSPSAGGEARVHVLVSSWRSGSFFVGOLFNOHP 64
DB 80 RQDITAFSQTALRLRNVTQLAGELGIAAP--BPRRHVLLMATRTGSSFFVGEFNOOG 136
QY 65 DVFYLMPEPAHWV--WTTLSQCSAATLHMV--RDLVRSVFLCDMDVFDAYL---PWRNL 117
DB 137 NIFYLFEPLMHIERTVTFEPGANAAGSALVYRDVLOQLLLCDLYLESFISPAPEEHLT 196
QY 118 SDFLOWAVSRALCSPACSAFPRGAISSEAVCKPFCARQSFFLAREACRSYSHVVLKEVR 177
DB 197 AALFRGSGSHSLCEPEVCTPSLKKVPEKYCKNRCGPNITLAAEACRRKQHMALKTVR 256
QY 178 FPNLQVLYPLSDPALNLRIHVLRDPRAVLRSEQTAKALARDNGIVLGTNGTW----V 233
DB 257 IRQLEFLQPLAEDPRLDLRIQLVRDPRAVLVSRVAFS-----GKYESNKKWAA 306
QY 234 EADPGLRV-----VREVCRSYHVRVIAEATLKPPFPLRGYRLVRFPEDLAREPLAIRALY 288
DB 307 EGAEAPLOEDEVORLKGNCES-IRLSAELGLRQPLWLRGYLMVRYEDVARAPLRKALEMY 365
QY 289 APTGLSLTPQLAEAWHITHGSGPGARREAFKTSRRNALNVSOAHRHALPFKIRRVQEL 348

Db 366 RFAGIHPTQVEEMIRANT--QAPQDSNGIYST-QKNSSEOFKWRFSIPFKLAQVQDA 422
 Qy 349 CAGALQLLQVPRVYSEQRNLADLV 375
 Db 423 CEPAMRLFGYKLASSAQELNRSLSLL 449

RESULT 5
 707229
 hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
 C:Accession: G70729
 R:Colet, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: G70729
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-388 <COL>
 A:Cross-references: UNIPROT:Q50695; GB:Z77163; GB:AL123456; MID:g3261610; PIDN:CAB00968.
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv2267c

Query Match 6.2%; Score 128; DB 2; Length 388;
 Best Local Similarity 22.9%; Pred. No. 0.002;
 Matches 81; Conservative 46; Mismatches 121; Indels 106; Gaps 18;

Qy 43 VLVLSSNRSGSFVQLP-----NQH-----PDVFLMEPAWVHTLSSQSAATLH 89
 Db 83 IFIVGHWRGTGTLHELLVDDRHTGPTGYECLAPHHFLITE-----WFAPVBFVLSKH 137
 Qy 90 MAVRDVRSVFLCDMDVFDVLPVRRNLSLDFQAVSRALCSPACSAFPGALSSSEAVC 149
 Db 138 RA-----MDNMD-----LSLHHQDEDFVCM-QGLPSPLYTIAFN----- 173
 Qy 150 KPLCAROSFTLARBACRSYSHWLKVEVFFNLQVLY-----PLLSDPALNLR----- 197
 Db 174 RPPQVEEVLDEQVAPRELE--INKRTLFRVQVYFRRKTVILKNPHTSFRIKVLLEV 231
 Qy 198 -----VHLVRDPRVLRSEQTAKALARDNGIVLGTNGTWVEADPGL--RVVREVCRS 249
 Db 232 FPQAKFIHVRDPVYVVPSTTHLKALYRIHGLQPT-----FDGLDDKVVSTYVDLY 284
 Qy 250 VRIAEATLKPPFLGRYRLVRPDLAREPLAEIRALYAFGLS-----LTPQLEAMHN 305
 Db 285 RKLDEGRELDVPT----RFYELRYEDLIGDPEGQLRLYQHLGLGDFECYLPRLRQVLY-- 338
 Qy 306 ITHGSGPGARAEFTSS-----RNALNVSQW-----RHALPFAKIR 343
 Db 339 -----ADHADIYKNTSYQLTVQORAI-VDEHWGEIIRYGYDRHTPEPARLR 383

RESULT 6
 E95934
 probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc
 C:Species: Sinorhizobium meliloti
 C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: E95934
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95934
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-307 <KUR>

A:Cross-references: UNIPROT:Q92VG4; GB:AL591985; PIDN:CAC49141.1; PID:g15140626; GSPDB:GB
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebault, P.; Vandebol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMB21237
 A:Genome: plasmid

Query Match 5.7%; Score 118; DB 2; Length 307;
 Best Local Similarity 28.1%; Pred. No. 0.011;
 Matches 50; Conservative 27; Mismatches 71; Indels 30; Gaps 9;

Qy 191 PALNLRIVHLVRDPRV-----LRSREQTAKALARDNGIVLGTNGTWVEADPGLRVREVC 246
 Db 149 PGIDMRVHLVRDGRGVAVWSLLKGYERDA-----SGLQKEIKP-KSVFRTAL 195
 Qy 247 R-SHVRIAE---AATLKPPFLGRYRLVRPDLAREPLAEIRALYAFGLSLTPQLEAW 302
 Db 196 RWSMVNLAVELYLSRKLGSSEKVMR-----VRYEDFASDPVAVMQQIGTFLELDLS-QVGT 249
 Qy 303 IHNITHSGPG--ARREAFKTSRNALNVSQWHRHALPFAKIRRVQELCAGALQLLY 358
 Db 250 LEN-GEAMGPGHQVAGNRLNRNLSIALNKDETRTRMPARQQVVSFQRLGGWMLRRYGY 306

RESULT 7
 B83229
 probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aerugin
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: B83229
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Labbig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathoge
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: B83229
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-523 <STO>
 A:Cross-references: UNIPROT:Q91626; GB:AE004451; GB:AE004091; MID:g9945958; PIDN:AAG0352
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA0136

Query Match 4.9%; Score 100.5; DB 2; Length 523;
 Best Local Similarity 26.8%; Pred. No. 0.8;
 Matches 38; Conservative 19; Mismatches 52; Indels 33; Gaps 6;

Qy 14 LLLAQTFLLFLYSRPGPSSPA-----CGEARVHVVLSSMRSGSFVQGLFNQHPDVFYL 69
 Db 385 LALAEIIRRFVAVKAPGAPARSLSGNLQKFTL-----GREILQAPRLVA 432
 Qy 70 MEPAWHVHTLSSQSAATLH---MAVRDLVRSVFLCDMDVFDVLPVRRNLSLDFQWAVS 126
 Db 433 AHPTWGV-----DVGAALILHIALALADAGTALILVWSEDLDELFL-----LSDRI----- 478
 Qy 127 RALCSPPACSAFPRGATSSAV 148
 Db 479 AALCSGRCLCPAVATASAPQKV 500

RESULT 8
 T03412
 malate synthase (EC 4.1.3.2) - maize

162 REACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPAVLRSREQTAKALARD 221
176 -----LWLPTIR-----RLYP-----NAKFLMIRDARAVHSMIERKVPVA-- 212
222 NGIVLGTN-----GTWVEADPGLRVVRECVSHVRIAEEATLKPPFFLRGRVLRVPEDL 276
213 -----GYNTSDEISMFWQNMQLKMTFQCNN-----APQOCIK-----VYERL 252
277 ARPLAERIALYAFGLSLTPLEAWIHNITHGSGPGARREAFKTS--RNALNVS--QAW 333
253 IQRPAESILITNPLDLPFSQMLR--HODLIGDEVLDNDQFESASOVKNSTKALTWS 310
334 RHALPFAKIRRVQELCAGALQLGY-----RPVYS 363
311 FDCFSBETLAKLDDV-APFLGILGYDTSISKPDYS 344

RESULT 10
E83023
3-deoxy-D-manno-octulosonic-acid (KDO) transferase PA4988 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: E83023
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Brady, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: E83023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-425 <STO>
A:Cross-references: UNIPROT:Q9HUH7; GB:AE004911; GB:AE004091; NID:g9951264; PIDN:AAG08373
A:Experimental source: strain PA01
C:Genetics:
A:Gene: waaA; PA4988
C:Superfamily: Chlamydia trachomatis 3-deoxy-manno-octulosonate cytidyltransferase

Query Match 4.5%; Score 91.5; DB 2; Length 425;
Best Local Similarity 22.2%; Pred. No. 3.8;
Matches 88; Conservative 41; Mismatches 141; Indels 127; Gaps 20;

31 PSSPAGCAEAVHVLVLSWSRSGSFVGLFQHPDV---FYLMEPAHWVTTLSQGSAA 87
44 PEVPPGG-IWHAIVSGESIAAPVWVALLERHPQLPVTVCTMT-----GS----- 90
88 LHMAVRLVRSVFLCDMDVDFDAYLPWRNLSDLFQWAVSRALCSPACSAFPRGAISSEA 147
91 -----ERIRALF--GQVRHCYLPY----DL-PWAAARFL-----DRVPRLAIVMET 131
148 VCKP-----LCARQSF-----TLAREACRYSYHV-----VLKEVRFNQL----- 182
132 ELWPNHITHACVRGIPVALANARLSERSARGYARFAGITRPLMLAELSWIAVQTEAEARF 191
183 -----VLYPLSDPALNLRIVHLVRDPAVLRSREQTAKALARDNGIVLG 227
192 RRLGARPECVSVTSGIKFDIRIDPOLPLAAALREEDATARPLWIAASTHAGEDIEVLA 251
228 TNGTWVEADPGLRVVRECVSHVRIAEEATLKPPFFLRGRVLRVPEDLAR-----E 279
252 AHRRLLETRPDALLI-----LVPRHPERFAGVHCLCRGEGFATVRRSGGE 296
280 PLA-----ETRALY-----AFTGLSLTPLEAWIHNITHGSGPG-----ARRA 318
297 PVARATQVLLGDTMGEFLFYALADIAFVGGSLVFN---GGHNLLEPAALGKPVFAGPHL 353
319 FKTSSRNALNVSQWRHALPFAKIRRVQELCAGALQL 355
354 F-----NFLDIAAQLRDAGALLEVTVDAGELCDGLARL 385

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.;
F.G.; Nunes, L.R.; Oliveira, M.A.P.; Matsukuma, A.Y.; Menck, C.P.M.; Miracca, E.C.; Miyaki, C.Y.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0933
C:Superfamily: ferrous iron transport protein B; translation elongation factor Tu homolog

Query Match 4.5%; Score 91.5; DB 2; Length 617;
Best Local Similarity 25.7%; Pred. No. 6.2;
Matches 47; Conservative 24; Mismatches 73; Indels 39; Gaps 7;

QY 11 VTALLAQAQTLLPLLVSRPQSSPAGGEARVHVLLVLSWRSGSSVGVQLPQNHQPDVFFYLM 70
DB 456 VGGIILALTLLWFLAPPEPPSGVGTGPAIYY-----SPAGRICHVMAGFFAPL 504
QY 71 EPAHWVWTTLSQSSNATLHMVRDLRSVFLC---DMDVFDAYLPWRNLSDLFO--WA 124
DB 505 GFSWQICIALIPGLAAR-EYAVSSLATVYALSANEEMAV-----RALSPLIQDGWS 555
QY 125 VSRALCSPPACSAFPRGAISSEAVCKPLCARQSFTLAREACRSYSHVVLKEVRFNQLVL 184
DB 556 LATAL-----SLLVWYIYAPMCISTLATIKRET-NSWKHMAMTAGYLFALAYL 602
QY 185 YPL 187
DB 603 AAL 605

RESULT 13
H86246
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H86246
R:Rheologs, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C.A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
Rizzo, M.; Rooney, T.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712

```

A;Accession: H86246
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-820 <STO>
A;Cross-references: UNIPROT:Q9SXB4; GB:AE005172; NID:G5734728; PIDN:AAD49993.1; GSPDB:GN000000000
C;Genetics:
A;Map position: 1
C;Superfamily: S-receptor kinase; protein kinase homology; S-locus-specific glycoprotein

Query Match          4.5%; Score 91.5; DB 2; Length 820;
Best Local Similarity 22.8%; Pred. No.8.9;
Matches 78; Conservative 38; Mismatches 91; Indels 135; Gaps 21;

Qy      2 WLP-----RVSSSTAVTALLAQTFLLFLVSRPSSPAGGEARVHVLVLS-----48
|||    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
157 WLPNMLVTGARIGCGNVT-----ITSWKSPPDPSPG-SYTAALVLAAYPELF 203

Db      49 -----WRSGSFVGOLFNOHPDVF-----YLMEPAWHVWTTLSCQSAAATLHM 90
|||    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
204 IMNNNNNSTVRRSG-PWNGQMENGLPDVIYACFLYRFIVNDDTNGSVTMYSANDSTLRY 262

Qy     91 AVRDLVRVSFLCDMDVFDAYLPW-----RRNLIS-----DLFQWAVSRALCSGP---PA 134
|||    :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

Db 263 FYMDYRGSVIRRD-----WSETRNTWTVGLQVPATECDNYRCGEFATCNPRKNPL 313
QY 135 CSAF-----PRGAI-----SSEAVCK-----PL-CARQSFLLAREACRSYSHVVLKEVRFNNL 161
Db 314 CSIRGFRPNLWNGNWSGGCTRRVPLOCRQNNNGS----- 353
QY 182 QVLYPLLSDPALNLRIVHL-----VRDPRAVLRSEQTAKALARDNGIVLGT---NG 230
Db 354 -----ADGFLRLRMKLPDFARRSEASEPEC-LRTCLQTCSCIAAAHGLGYGCHWING 405
QY 231 TWVEADPLRVREVCRS-----HVRIA--EAATLKPPFPLRG 266
Db 406 SLVDS-----QELSGAGLDLYRLAHSEIKTKDRPILIG 440

RESULT 14
A75328
PhoH-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: A75328
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: A75328
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-380 <WHI>
A:Cross-references: UNIPROT:Q9RSY1; GB:AE002038; GB:AE00513; NID:G6459790; PIDN:AAF1154
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1988
A:Map position: 1
C:Superfamily: phosphate starvation-induced protein

Query Match 4.4%; Score 91; DB 2; Length 380;
Best Local Similarity 26.5%; Pred. No. 3.7;
Matches 68; Conservative 25; Mismatches 104; Indels 60; Gaps 14;

QY 124 AVSRALCSPACGAFPRGAISSSEAVCKPLCARQSFLLARE-----ACRSYSHVVLKEVRF 179
Db 2 ALSECLFSP-----FPRPLGED--CAPAGARTGDKLTDQTNIAPEADQTVTATVLN 55
QY 180 NLQVLYPLLSDPALNLRIVHLVRDPRAVLRSEBOT-AKALARDNGIVLGTNGTWVE-ADP 237
Db 56 DQREAYALLGANDANLR-----RMRELTRAKLIARTGETVTITGDAADVEGAER 103
QY 238 GLRVREVCRSHVRIAEATLKPPPLRGVRLVRFEDLAREPLAEIRALYFTGLSL-- 295
Db 104 MVRDALDVRS-----GGELTPDLSLRS-----ARLSSEGRSLAAETQV-----NGLTLP 149
QY 296 -----TPQLEAWI-----HNITHSGPGARREAFKTSRNALNVSOAWRHALPFKIRRV 345
Db 150 GLRKPTFGQKLYLDLINESDITFGVPAG---TCKTYMAVAMAV-----QALAKKKVKRI 201
QY 346 --QELCAGALQLGYRP 360
Db 202 ILTRPAVEAGEKLGFLP 218

RESULT 15
E75611
glucosamine-fructose-6-phosphate aminotransferase-related protein - Deinococcus radiodur
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: E75611
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: E75611
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <WHI>
A:Cross-references: UNIPROT:Q9RY29; GB:AE001862; GB:AE001825; NID:G6460468; PIDN:AAF1221
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0154
A:Map position: 2

Query Match 4.4%; Score 90.5; DB 2; Length 349;
Best Local Similarity 24.1%; Pred. No. 3.6;
Matches 78; Conservative 34; Mismatches 133; Indels 79; Gaps 15;

QY 33 SPAGGEARVHVLVLSWSRGSSVFQGFNOHPDVFYLM-----EPAM-----H 75
Db 2 TPDASAAPAEPLMLREMRAPAVVRELLNENQDAVARLANAIRERQPAVAVTTIARGSSDH 61
QY 76 VMTTLSQGSAAATLHMAVRDLVRSVFLCDMDVFDAYLPMWRNLSDLFQWAVSRALCSPPPAC 135
Db 62 ACTVLKYALETQSLPVASLGPSVH-----TLYGA-----RUDLAGALVIAVSQSGASPDVV 113
QY 136 S-----AFPRGAISSEAV---CKPLCARQSFLL-----AREACRSYSHVVLKEVRF 180
Db 114 ENVRMAREGALTVALVNVNVEDSLAEAEFTLPLRCGPEKAVAATKSY---LASLCAF- 168
QY 181 LQVLYPLLSDPALNLRIVHLVRDPRAVLRSEBOTAKALA-----RNGIVLGTNGTWVEA 235
Db 169 LPVLAELTGDALK-RALDALPEQLAHTLTLEDAAHAEAGRYFADNLLTLARGHYGV 227
QY 236 DPGLRVREVCRSHVRIAEATLKPPPLRGVRLV-----RPE 274
Db 228 QEAALKKKTCTGHIHAEYSAA-----EFSGPKRLLAEGLPLIGFASADAWEATQAYD 282
QY 275 DLAREPLAEIRALYFTGLSL-TP 297
Db 283 DL-RAAGADLRLLGPVAGADLPTP 305

Search completed: June 23, 2005, 08:50:17
Job time : 10.9593 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 43.3472 Seconds
(without alignments)

4666.302 Million cell updates/sec

Title: US-10-697-828-13

Perfect score: 2056

Sequence: 1 MWLPRVSSTAVTALLAQTF.....LPRGLNGFTWASSTASHPRN 395

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2056	100.0	395	2 Q9GZX3	Q9GZX3 homo sapien
2	1729.5	84.1	390	2 Q7LCN3	Q7LCN3 homo sapien
3	1729.5	84.1	411	2 Q9GZS9	Q9GZS9 homo sapien
4	1708	83.1	395	2 Q9QUF4	Q9QUF4 mus musculus
5	1177	57.2	392	2 Q6GN39	Q6GN39 xenopus lae
6	1024	49.8	386	2 Q8NC05	Q8NC05 homo sapien
7	1019	49.6	386	2 Q9Y5R3	Q9Y5R3 homo sapien
8	1012	49.2	370	2 Q8IV46	Q8IV46 homo sapien
9	984	47.9	388	2 Q8RI11	Q8RI11 mus musculus
10	977	47.5	388	2 Q9W0E5	Q9W0E5 mus musculus
11	663	32.2	483	2 Q794G9	Q794G9 mus musculus
12	663	32.2	530	2 Q82T76	Q82T76 mus musculus
13	660	32.1	530	2 Q80WV3	Q80WV3 mus musculus
14	657.5	32.0	483	2 Q9UED5	Q9UED5 homo sapien
15	657.5	32.0	530	2 Q9Y4C5	Q9Y4C5 homo sapien
16	622.5	30.3	484	2 Q9EP78	Q9EP78 m n-acetyl
17	619.5	30.1	484	2 Q99NR0	Q99NR0 mus musculus
18	615	29.9	486	2 Q75667	Q75667 homo sapien
19	614	29.9	486	2 Q9NS84	Q9NS84 homo sapien
20	593	28.8	420	2 Q6DBY9	Q6DBY9 brachydanio
21	593	28.8	485	2 Q8XQ88	Q8XQ88 rattus norv
22	578.5	28.1	479	2 Q7LGC8	Q7LGC8 homo sapien
23	576.5	28.0	479	2 Q75099	Q75099 homo sapien
24	569	27.7	472	2 Q88199	Q88199 mus musculus
25	564.5	27.5	411	2 Q9EQC0	Q9EQC0 mus musculus
26	558.5	27.0	411	2 Q43916	Q43916 homo sapien
27	555	27.0	474	2 Q9QZ12	Q9QZ12 rattus norv
28	532.5	25.9	458	1 C6ST_CHICK	Q92179 gallus gall
29	463.5	22.5	441	2 Q93403	Q93403 torpedo cal
30	393.5	19.1	257	2 Q79415	Q79415 mus musculus
31	347.5	16.9	225	2 Q6RY62	Q6RY62 cavia porce

32 288.5 14.0 304 2 Q677T4
33 270.5 13.2 486 2 Q9VMC3
34 270 13.1 345 2 Q7Q1F9
35 193 9.4 120 2 Q9DOK5
36 190 9.2 119 2 Q95JA8
37 169.5 8.2 363 2 Q9VMC4
38 163.5 8.0 363 2 Q8MZD1
39 133 6.5 303 2 Q93JB6
40 128 6.2 388 1 YM67 MYCTU
41 128 6.2 388 1 YM90 MYCBO
42 123.5 6.0 1222 2 Q81ZU8
43 118 5.7 307 2 Q92VG4
44 114.5 5.6 274 2 Q8PRA0
45 113.5 5.5 496 2 Q9EX08

ALIGNMENTS

RESULT 1

Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal N-
DE acetylglucosamine-6-O-sulfotransferase) (Carbohydrate N-
DE acetylglucosamine 6-O) sulfotransferase 6).
GN Name=GST4beta; Synonyms=CHST6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096027; PubMed=1181564; DOI=10.1093/glycob/11.1.1.75;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization of the galactose/N-
RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
RT family";
RL Glycobiology 11:75-87(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Axama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene";
RL Nat. Genet. 26:237-241(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Basak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smallos D.E., Schnerch A., Schein J.E.,

```

RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Director MGC Project;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF280086; AAC48444.1; -.
DR EMBL; AF219991; AAG26327.1; -.
DR EMBL; BC074883; AAH74883.1; -.
DR EMBL; BC074834; AAH74834.1; -.
DR EMBL; AF219990; AAC26325.1; -.
DR Genew; HGNC:6938; CHST6.
DR GO; GO:0005794; C:Golgi apparatus; TAS.
DR GO; GO:0001517; F:N-acetylglucosamine 6-O-sulfotransferase ac...; TAS.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; IC.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

Query Match 100.0%; Score 2056; DB 2; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.6e-165;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVQQLF 60
DB 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVQQLF 60
QY 61 NQHPDVFYLMPEAWHVVTTLSQSSAATLHMVRLVRSVFLCDMDVDFAYLPWRNLSL 120
DB 61 NQHPDVFYLMPEAWHVVTTLSQSSAATLHMVRLVRSVFLCDMDVDFAYLPWRNLSL 120
QY 121 FQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 180
DB 121 FQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 180
QY 181 LQVLYPLSDPALNLRVHVRDPRVAVLRSEOTAKALARDNGIVLGTNGTWVADPGLR 240
DB 181 LQVLYPLSDPALNLRVHVRDPRVAVLRSEOTAKALARDNGIVLGTNGTWVADPGLR 240
QY 241 VREVCCHSVRIAEATLKPPPLRGYRLVRPDLAREPLAEIRALYAFGLTLPQLE 300
DB 241 VREVCCHSVRIAEATLKPPPLRGYRLVRPDLAREPLAEIRALYAFGLTLPQLE 300
QY 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFKIRRVQELCAGALQLLGYRP 360
DB 301 AWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFKIRRVQELCAGALQLLGYRP 360
QY 361 VYSEDEQRNALDLVLRGLNGFTWASSTASHPRN 395
DB 361 VYSEDEQRNALDLVLRGLNGFTWASSTASHPRN 395

RESULT 2
Q7LCN3 PRELIMINARY; PRT; 390 AA.
AC Q7LCN3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Name=I-GlcNAc-6-ST;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=99423499; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;

```

```

RA Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RT sulfotransferase that is highly restricted to intestinal tissue.";
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF176839; AAD56001.1; -.
DR EMBL; AF176838; AAD56000.1; -.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 390 AA; 43832 MW; 132D1C3B6BE38179 CRC64;

Query Match 84.1%; Score 1729.5; DB 2; Length 390;
Best Local Similarity 85.8%; Pred. No. 7e-138;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVQQL 59
DB 1 MWLPRVSTAVTALLAQTFLLFLVSRPGSPAGGEARVHVLVLSWSRSGSFVQQL 60
QY 60 NQHPDVFYLMPEAWHVVTTLSQSSAATLHMVRLVRSVFLCDMDVDFAYLPWRNLSL 119
DB 61 FQHPDVFYLMPEAWHVVTTLSQSSAATLHMVRLVRSVFLCDMDVDFAYLPWRNLSA 120
QY 120 LQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 179
DB 121 FQWAVSRALCSPACSAFPRGAISSBAVKPLCARQSFLLAREACRSYSHVVLKEVRFF 180
QY 180 NQVLYPLSDPALNLRVHVRDPRVAVLRSEOTAKALARDNGIVLGTNGTWVADPGL 239
DB 181 NQVLYPLSDPALNLRVHVRDPRVAVLRSEOTAKALARDNGIVLGTNGTWVADPGL 240
QY 240 RVVRCCHSVRIAEATLKPPPLRGYRLVRPDLAREPLAEIRALYAFGLTLPQL 299
DB 241 RLIREVCCHSVRIAEATLKPPPLRGYRLVRPDLAREPLAEIRALYAFGLTLPQL 300
QY 300 EAWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFKIRRVQELCAGALQLLGYR 359
DB 301 EAWIHNITHGSGPGARREAFKTSRNALNVSQAWRHALPFKIRRVQELCAGALQLLGYR 360
QY 360 VYSEDEQRNALDLVLRGLNGFTWAS 387
DB 361 VYSAQQQRDLTLVLRPGDPHFSWAS 388

RESULT 3
Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9; Q9UB3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Intestinal GlcNAc-6-sulfotransferase (Intestinal N-acetylglucosamine-
DE 6-O-sulfotransferase).
GN Name=CHST5;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086; DOI=10.1038/79987;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RX "Macular corneal dystrophy type I and type II are caused by distinct

```

```

RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR ENBL; AF246718; AAC28023.1; -.
DR ENBL; AF219991; AAC26326.1; -.
DR Genew; HGNC:1973; CHST5.
DR GO; GO:0008146; F:sulphotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match      84.1%; Score 1729.5; DB 2; Length 411;
Best Local Similarity 85.8%; Pred. No. 7.4e-138;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPVSVSTAVTALLAQ-TFLLLVSRPGSPAGGEARVHVLVLSRSGSSFGQL 59
DB 22 MWLPVSVSTAVTALLAQ-TFLLLVSRPGSPAGGEDRVHVLVLSRSGSSFGQL 81
QY 60 FNOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLPMRRNLSD 119
DB 82 FSQHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLNRSIFLCDMDVDFDAYMPCSRNLSA 141
QY 120 LFQWAVSRALCSPACSAFPRGAISSEAVCKPLCAROSFTLAREACRSYSHVVLKEVRFF 179
DB 142 FFWWATSRALCSPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRFF 201
QY 180 NLQVLYPLSDPALNLRIVHVRDPRVAVLSREQTAKALARDNGIVLGTNGTWVEADPGL 239
DB 202 NLQVLYPLSDPALNLRIVHVRDPRVAVLSREAAAGPILARDNGIVLGTNGKWEADPHL 261
QY 240 RVREVCRSVHRIAEEATLKPPLGRYLRVRFEDLAREPLAEIRALYAFGLSLTPQL 299
DB 262 RLREVCRSVHRIAEEATLKPPLGRYLRVRFEDLAREPLAEIRALYAFGLSLTPQL 321
QY 300 EAWIHNTHSGGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLGYR 359
DB 322 EAWIHNTHSGGPGARREAFKTSRNALNVSQAWRHALPFTKILRVQEVCAQALQLGYR 381
QY 360 PVYSEDEQRNALDLVLPRLNGFTWAS 387
DB 382 PVYSAQQRDLTDLVLPRLGPDHFSWAS 409

RESULT 4
Q9QUP4 PRELIMINARY; PRT; 395 AA.
AC Q9QUP4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 03-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase.
GN Name=Chst5; Synonyms=I-GlcNAC-6-ST;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57Bl/6; TISSUE=Intestine;
RP STRAIN=C57Bl/6; PubMed=10491328; DOI=10.1006/bbrc.1999.1324;
RX MEDLINE=99423499; PubMed=10491328;
RA Lee J.K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and characterization of a mammalian N-acetylglucosamine-6-
RL sulfotransferase that is highly restricted to intestinal tissue.";
RN Biochem. Biophys. Res. Commun. 263:543-549(1999).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=C57Bl/6; TISSUE=Intestine;
RC STRAIN=C57Bl/6; Rosen S.D., Hemmerich S.;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF176841; AAD56003.1; -.
DR ENBL; AF176840; AAD56002.1; -.

```

```

DR MGD; MGI:1931825; Chst5.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase.
KW Transferase.
SQ SEQUENCE 395 AA; 44537 MW; 3FDF71843ED3D38E CRC64;

Query Match      83.1%; Score 1708; DB 2; Length 395;
Best Local Similarity 83.0%; Pred. No. 4.7e-136;
Matches 328; Conservative 24; Mismatches 43; Indels 0; Gaps 0;

QY 1 MWLPVSVSTAVTALLAQ-TFLLLVSRPGSPAGGEARVHVLVLSRSGSSFGQLF 60
DB 1 MRLPREFSSTVMLSLMVQTGILVLSRQVPSFAGLGERVHVLVLSRSGSSFGQLF 60
QY 61 NOHPDVFLMEPAWHVWTTLSQGSAAATLHMAVRDLVRSVFLCDMDVDFDAYLPMRRNLSD 120
DB 61 SQHPDVFLMEPAWHVWTTLSQGSAPALHMAVRDLRSVFLCDMDVDFDAYLPMRRNLSD 120
QY 121 FQWAVSRALCSPACSAFPRGAISSEAVCKPLCAROSFTLAREACRSYSHVVLKEVRFF 180
DB 121 FQWAVSRALCSPVCEAFAGNISSEVCKPLCATRPFGLAQEACSSYSHVVLKEVRFF 180
QY 181 LOVLYPLSDPALNLRIVHVRDPRVAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
DB 181 LOVLYPLSDPALNLRIVHVRDPRVAVLSREQTAKALARDNGIVLGTNGTWVEADPGLR 240
QY 241 VREVCRSVHRIAEEATLKPPLGRYLRVRFEDLAREPLAEIRALYAFGLSLTPQL 300
DB 241 VNEVCRSVHRIAEEALHKKPPPLQDRYLRVRFEDLARDPLTVIRELYAFGLSLTPQL 300
QY 301 AWIHNTHSGGPGARREAFKTSRNALNVSQAWRHALPFAKIRRVQELCAGALQLGYR 360
DB 301 TWIHNTHSGGPGARREAFKTSRDALSVSQAWRHLPFAKIRRVQELCAGALQLGYR 360
QY 361 VYSEDEQRNALDLVLPRLNGFTWASSTASHPRN 395
DB 361 VHSLEQRDLSDLLPRGMDSPKASSTEKQEPES 395

RESULT 5
Q6GN39 PRELIMINARY; PRT; 392 AA.
AC Q6GN39;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE MGC83048 protein.
GN Name=MGC83048;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
SEQUENCE FROM N.A.
RP TISSUE=Embryo.
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haiech F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Roha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

```


RA Krzyzowski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.;
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Embryo;
 RA Klein S., Gerhard D.S.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC073681; AAH73681.1; -;
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 SQ SEQUENCE 392 AA; 45969 MW; 7712810F8B4704D7 CRC64;
 Query Match 57.2%; Score 1177; DB 2; Length 392;
 Best Local Similarity 57.4%; Pred. No. 4.2e-91;
 Matches 220; Conservative 68; Mismatches 91; Indels 4; Gaps 3;
 QY 5 RVSTAVTALLAQTFLLLFLVRPG--PSSPAGGEARVHVLVLSWRSGSFVQLFNQ 62
 DB 5 RPNVTVAGFFLQTLFLIYSRHTVLPDTETKTE-KVHLLLSWRSGSFVQLFNQ 63
 QY 63 HPDVFYLMPEPAHVHTTSLQGSAAATLHMAVDRDLVRSVFLCDMDVDFAYLFWRNLSDLFQ 122
 DB 64 HPDVFYLMPEPAHVHTTSLQGSAAATLHMAVDRDLVRSVFLCDMDVDFAYLFWRNLSDLFQ 123
 QY 123 WAYSRALCSPACAPRGAISSEAVCKPCARQSFLLAREACRSYSHVVLKEVRFNQ 182
 DB 124 WAYSRALCSPACAPRGAISSEAVCKPCARQSFLLAREACRSYSHVVLKEVRFNQ 183
 QY 183 VLYPLSDPALNLRIVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPLRVV 242
 DB 184 VLYPLSDPALNLRIVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPLRVV 242
 QY 243 REVCRSHVRIABATLPPFLGRYLVRFEDLAREPLAEIRALYFTGLSLTPQLEAW 302
 DB 243 REICQSHAQMYETAMDKAPSLKGRYLVRYEDVDRPLREINQMVEFANLKTAKLKNW 302
 QY 303 IHNTHGSGPGARREAFKTSRRNALVNSQAWRHALPPAKIRRVQELCAGALQLLGYRPVY 362
 DB 303 FYNIHGVGPGTKKBEFTTSRRNALVNSQAWRHALPPAKIRRVQELCAGALQLLGYRPVY 362
 QY 363 SEDEQRNALDLVLPRLNGFTW 385
 DB 363 SEXERKDSMDVLPKRNQFSW 385
 RESULT 6
 Q8NCG5 PRELIMINARY; PRT; 386 AA.
 AC Q8NCG5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (Human).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,

RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
 RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK074746; BAC11177.1; -;
 DR GenBank; HGNC:1972; CHST4.
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer_1; 1.
 KW Transferase.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;
 Query Match 49.8%; Score 1024; DB 2; Length 386;
 Best Local Similarity 54.8%; Pred. No. 3.7e-78;
 Matches 217; Conservative 44; Mismatches 107; Indels 28; Gaps 7;
 QY 14 LLLAQTFLLLFLVSRP-----GSSPAGGEARVHVLVLSWRSGSFVQL 59
 DB 2 LLPKMKLLLLFLVSQWAILALFPHMYSHNISLSMKAQPERHVLVLSWRSGSFVQL 61
 QY 60 FNOHPDVFYLMPEPAHVHTTSLQGSAAATLHMAVDRDLVRSVFLCDMDVDFAYLFWRNLS 118
 DB 62 FQHPDVFYLMPEPAHVHTTSLQGSAAATLHMAVDRDLVRSVFLCDMDVDFAYLFWRNLS 121
 QY 119 DLFQWAVSALCSPACAPRGAISSEAVCKPCARQSFLLAREACRSYSHVVLKEVRFNQ 178
 DB 122 SUFQWENSALCSPACAPRGAISSEAVCKPCARQSFLLAREACRSYSHVVLKEVRFNQ 181
 QY 179 FNLQVLYPLSDPALNLRIVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPLRVV 237
 DB 182 FNLQVLYPLSDPALNLRIVHVRDPAVLRSEQTAKALARDNGIVLGTNGTWVEADPLRVV 241
 QY 238 GLRVVRECRSHVRIABATLPPFLGRYLVRFEDLAREPLAEIRALYFTGLSLTPQLEAW 296
 DB 242 PYYVMQVICSQLEIYK--TIQSLPKALQERYLVRYEDLAREPLAEIRALYFTGLSLTPQLEAW 299
 QY 297 POLEAWIHNTHGSGPGARREAFKTSRRNALVNSQAWRHALPPAKIRRVQELCAGALQLL 356
 DB 300 PHLQTVHNTKGMG--DHAFHTNARDNALVNSQAWRHALPPAKIRRVQELCAGALQLL 357
 QY 357 GYRPVYSEDEQRNALDLVLPRLNGFTWASSTASH 392
 DB 358 GYRHVSEDEQRNALDLVLPRLNGFTWASSTASH 392
 RESULT 7
 Q9V5R3 PRELIMINARY; PRT; 386 AA.
 ID Q9V5R3;
 AC Q9V5R3;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
 DE sulfotransferase GST-3).
 GN Name=GST3;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Tonsil;
 RX MEDLINE=99264336; PubMed=10330415; DOI=10.1083/jcb.145.4.899;
 RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
 RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
 RT "Sulfotransferases of two specificities function in the reconstitution
 RT of high endothelial cell ligands for L-selectin.";
 RL J. Cell Biol. 145:899-910(1999).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Tonsil;

RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.175;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.,
 RT "Chromosomal localization and genomic organization of the galactose/N-
 RT acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
 RL family";
 RN Glycobiology 11:75-87(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21325292; PubMed=11439191; DOI=10.1016/S0092-8674(01)00394-4;
 RA Yen J.-C., Hiraoka N., Pecyniak B., Nakayama J., Elies L.G.,
 RA Rabuka D., Hindegaal O., March J.D., Lowe J.B., Fakuda M.,
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RL Core1 extension beta 1,3-N-acetylglucosaminyltransferase";
 RN Cell 105:957-969(2001).
 DR EMBL; AF131235; AA033015.1; -
 DR EMBL; AF280088; AA484246.1; -
 DR EMBL; AF149783; AA48417.1; -
 DR GO; GO:0008146; F:sulfotransferase activity; TAS.
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR GO; GO:0006928; P:cell motility; TAS.
 DR GO; GO:0007267; P:cell-cell signaling; TAS.
 DR GO; GO:0006955; P:immune response; TAS.
 DR GO; GO:0006477; P:protein amino acid sulfation; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 KW Selectin; Transferrase.
 SQ SEQUENCE 386 AA; 45133 MW; OC3BB4022417143A CRC64;

 Query Match 49.6%; Score 1019; DB 2; Length 386;
 Best Local Similarity 54.5%; Pred. No. 9.7e-78;
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

 Qy 14 LLLAQTLLFLVSRP-----GPSSPAGGEARVHVVLVLSWRSGSFVQGL 59
 Db 2 LLPKMKLLFLVSRP-----GPSSPAGGEARVHVVLVLSWRSGSFVQGL 61

 Qy 60 FNOHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLVRSFLCDMDVFDAYL-PWRNLS 118
 Db 62 FGQHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLVRSFLCDMDVFDAYLMEGPRRQS 121

 Qy 119 DLFQWAVSRALCSPGAPGAISSAEVCKPCARQSFTLAREACRSYSHVVLKEVRF 178
 Db 122 SLFQWNSRALCSPACDIIPODEIIIPRAHCRLLCSQOPPEVVEKACRSYSHVVLKEVRF 181

 Qy 179 FNLQVLYPLSDPALNRIHVLRDPRVLRSEQTAKALARDNGIVLGTNGTWV-BADP 237
 Db 182 FNLQSLYPLKPSLNLHIVLRDPRVLRSEQTAKALARDNGIVLGTNGTWV-BADP 241

 Qy 238 GLRVVRECRSHVIAEAATLKP-PPFLRGYRLVREEDLAREPLAEIRALYFTGLSLT 296
 Db 242 PYYVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDLARAPVQTSRYEYFVGLLEF 299

 Qy 297 POLBAWHNTHGSGPGARREAFKTSRNALNVQAWRHLPFAKIRRVQSLCAGALQLL 356
 Db 300 PHLTQVHNITRGKMG--DHAFHTNARDALNVQAWRSLPYEKVSRQLQKACGDAMNLL 357

 Qy 357 GYRPVYGEQNRNALDLVLRGLNGFTWASSTASH 392
 Db 358 GYRHRVSEQQRNLLDLDL-----STWTVPEQIH 386

 RESULT 8
 Q8IV46
 ID Q8IV46 PRELIMINARY; PRT; 370 AA.
 AC Q8IV46;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE CHST4 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schaefer C.F., Bhat N.K.,
 RA Altschul S.P., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Plange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC035282; AAH35282.1; -
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

 Query Match 49.2%; Score 1012; DB 2; Length 370;
 Best Local Similarity 57.9%; Pred. No. 3.6e-77;
 Matches 206; Conservative 43; Mismatches 93; Indels 14; Gaps 6;

 Qy 40 RVHVLVLSWRSGSFVQGLFNOHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLVRSV 99
 Db 26 RVHVLVLSWRSGSFVQGLFNOHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLVRSV 85

 Qy 100 FLCDMDVFDAYL-PWRNLSDLFQWAVSRALCSPGAPGAISSAEVCKPCARQS 158
 Db 86 FLCDMSVFDAYLMEGPRRQSFLFQWNSRALCSPACDIIPODEIIIPRAHCRLLCSQOPF 145

 Qy 159 TLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNRIHVLRDPRVLRSEQTAKAL 218
 Db 146 EVVEKACRSYSHVVLKEVRFNQLVLYPLSDPALNRIHVLRDPRVLRSEQTAKAL 205

 Qy 219 ARDNGIVLGTNGTWV-EADPGLRVVRECRSHVIAEAATLKP-PPFLRGYRLVREEDL 276
 Db 206 MIDSRIVMGQHEQKLKEDQPYVVMQVICSQLEIYK--TIQSLPKALQERYLLVRYEDL 263

 Qy 277 AREPLAEIRALYFTGLSLTPQLEAMHNTTHGSGPGARREAFKTSRNALNVQAWRHA 336
 Db 264 ARAPVQTSRYEYFVGLLEFPLHQTQVHNITRGKMG--DHAFHTNARDALNVQAWRWS 321

 Qy 337 LPFAKIRRVQSLCAGALQLLQYRPVYSEDEQRNALDLVLRGLNGFTWASSTASH 392
 Db 322 LPYKVSRLQKACGDAMNLLGYRHRVSEQQRNLLDLDL-----STWTVPEQIH 370

 RESULT 9
 Q9RI11
 ID Q9RI11 PRELIMINARY; PRT; 388 AA.
 AC Q9RI11;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE L-selectin ligand sulfotransferase (Chst4 protein).

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
Genome Res. 10:1757-1771(2000).
[7]
RN SEQUENCE FROM N.A.
RP STRAIN=CS7BL/6J; TISSUE=Tongue;
RC Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF131236; AAD33016.1; -
DR EMBL; AK009113; BAB26078.1; -
DR MGD; MGI:1349479; Chst4.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
DR Transferase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 47.5%; Score 977; DB 2; Length 388;
Best Local Similarity 53.2%; Pred. No. 3.5e-74;
Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;

QY 14 LLLAQTEFLLEFLVS-----RPGSSPAGGEARVHVVLSSW 49
DB 2 MLLKKGLLMFLGQVIVVLFHMSVHRHLSQRESRP-----VHVVLSSW 50
QY 50 RSGSSFGVQLNQHDPVFLMEPAHWVTTLSQSSAATLHMAVDLRSVFLCDMDVFA 109
DB 51 RSGSSFGVQLGQHPDVFYLMPEPAHWVMTPTSSTAWKLHMAVDLRSVFLCDMSVFA 110
QY 110 YL-PWRNLSDLFQWAVSRALCSPACAPRGAISSSEAVCKPCARQSFLLAREACRSY 168
DB 111 YMNPGRKQSSLFQWESQRLSCAPVCDFFPAHEISSPKHCKLGGQQPDMVSEKCRSH 170
QY 169 SHVVLKEVRFFNLQVLPPLSDPALNLRIVHLVRDPAVLRSRQTKALARDNGIVLGT 228
DB 171 GFVVLKEVRFLSLQALYPLLTDPSLNLHVHLVRDPAVFRSREHTTIELMVDSHVILGQ 230
QY 229 N-GTWVEADPGLRVVREVCVRSHVRIAE-AFLKPPPLRGRYRLVRFPDLAREPLAEIRA 286
DB 231 HLETIKEEDQPYAMKICKSQVDIVKAIQTL--PEALQRYFLRYEDLVRAPLAQTR 288
QY 287 LYATGLSLTQLEAWIHNTGSGPGARRAFKTSRNALNVQAWRHALPFAKIRRVQ 346
DB 289 LYKFGVGLDFLPHLQTVVYVTVRGGMG--QHAFTNARNALNVQAWRWSLPYEKVSQ 346
QY 347 ELCAGALLQLGVRPVYSEDEQNLALDLV 375
DB 347 DACGENDLLGLVQRVSQEQGNLSLDLI 375

RESULT 11
Q794G9 PRELIMINARY; PRT; 483 AA.
AC Q794G9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (N-Acetylglucosamine 6-O-

DE Sulfotransferase-1).
GN Name=Onset-1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muranatsu T.,
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
sulfotransferase.";
J. Biol. Chem. 273:22577-22583(1998).
[2]
RN SEQUENCE FROM N.A.
RP Uchimura K., Kadomatsu K., El-Pasakhany F.M., Singer M.S., Izawa M.,
RA Kannagi R., Takeda N., Rosen S.D., Muramatsu T.,
RT "N-Acetylglucosamine 6-O-Sulfotransferase-1 Regulates Expression of L-
Selectin Ligands and Lymphocyte Homing.";
J. Biol. Chem. 279:35001-35008(2004).
RL EMBL; AB011452; BAA32139.1; -
DR EMBL; AB125058; BAA16775.1; -
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; F:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
DR Transferase.
SQ SEQUENCE 483 AA; 52830 MW; 831FA08F5FBD70E CRC64;

Query Match 32.2%; Score 663; DB 2; Length 483;
Best Local Similarity 39.5%; Pred. No. 1.8e-47;
Matches 146; Conservative 67; Mismatches 123; Indels 34; Gaps 9;

QY 36 GGEARVHVVLSSWSRSGSFGVQLFNQHPDVFYLMPEPAHWVTTLSQSSAATLHMAVDL 95
DB 113 GGDKRLVYVFTTWRSQSSFFGELFNQNPVEVFLYEPVHVWQKLYPGDAVSLQGAARDM 172
QY 96 VRSVFLCDMDVFPAYLP---WRNLSDL---FQWAVSRALCSPACAPRGAISS--SEAV 148
DB 173 LSALYRQCDLSVFLQYSPAGSGRNLTLTGFGAATNKCVCSSPLCPAYRKEVGLVDRV 232
QY 149 CKPLCARQSFLLAREACRSYSHVVLKEVRFFNLQVLPPLSDPALNLRIVHLVRDPAVL 208
DB 233 CKK-CPQRLAREECCRKYRTVVIKGRVDFVAVLAPLKKDPAIDLVKVIHLVRDPAVA 291
QY 209 RSREQTKALARDN-----GIVLGT--NGTWVEAD--PGLRVVREV 245
DB 292 SSRIRSHGLIRSLQVVRSDPRAHRMPPLEAAGHKGAKGCGMPADYHALGAMEVI 351
QY 246 CRSHVRIAEAAATLKPPPLRGRYRLVRFPDLAREPLAEIRALYFTGLSLTPQLEAWIHN 305
DB 352 CNSMAKTLQTA-LQPPDWLQCHYLVRVYEDLVGDPVKTLRRVYDFVGLLVSPENEQFALN 410
QY 306 ITHSGPGARRAFKTSRNALNVQAWRHALPFAKIRRVQELCAGALQLGVRPVYSED 365
DB 411 MTSGSGSSSK--PPVWSARNATQAANAWRTALTFOQIKQVEEFCYQPMVAVLGYRVSNSPE 468
QY 366 EORNALDLV 375
DB 469 EVDLSKTL 478

RESULT 12
O88276 PRELIMINARY; PRT; 530 AA.
ID O88276
AC O88276;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE N-acetylglucosamine-6-O-sulfotransferase long form.
GN Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryos;
RX MEDLINE=98380482; PubMed=9712885; DOI=10.1074/jbc.273.35.22577;
RA Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT sulfotransferase.";
RL J. Biol. Chem. 273:22577-22583 (1998).
DR EMBL; AB011452; BAA32138.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; P:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C363EC CRC64;

Query Match 32.2%; Score 663; DB 2; Length 530;
Best Local Similarity 39.5%; Pred. No. 2e-47;
Matches 146; Conservative 67; Mismatches 123; Indels 34; Gaps 9;

Qy 36 GGEARVHLVLSGWSGSSFGQLFNQHPDVFYLMPEAHVWTTLSQGSAAATLHMVRDL 95
Db 160 GGDKRQLVYVFTTWRSWGSSFFGELFNQNPVEFFLYEPVHVWQKLYPGDAVSLQGAARDM 219
Qy 96 VRSVFLCDMDVFAYLP---WRNLSDL---FQWAVSRALCSPACSAFPRGALS--SEAV 148
Db 220 LSALYRCDLSVFLQYSPAGSGGRNLTTLGFGAATNKVVCSSPLCPAYRKEVVGLVDDRV 279
Qy 149 CKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPRAYL 208
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLKDPALDLKVHLVRDPRAVA 338
Qy 209 RSREQTAKALARDN-----GIVLGT--NGTWVEAD-PGLRVVREV 245
Db 339 SSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLGAKGEGMGADYHALGAMEVI 398
Qy 246 CRSHVRIAEATLKPPPLRGRLVRFEDLAREPLAEIRALYAFGLSLTPQLEAWIHN 305
Db 399 CNSMAKTLQTA-LQPPDLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
Qy 306 ITHGSGPGARREAFKTSRRNALNVSOAHRHALPFKIRRVQELCAGALQLLGYRVPVSED 365
Db 458 MTSGSGSSSK--PFVVSARNATQANANWRTALTFOQIKQVEEFCYQPMVILGYRVNSPE 515
Qy 366 EORNALDLV 375
Db 516 EVKDLSTLL 525

RESULT 13
Q80WV3 PRELIMINARY; PRT; 530 AA.
AC Q80WV3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chst2 protein.
OS Name=Chst2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Moore T.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC051963; AAHS1963.1; -.
DR MGD; MGI:1891160; Chst2.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0008146; P:sulfotransferase activity; IDA.
DR GO; GO:0006044; P:N-acetylglucosamine metabolism; TAS.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
SQ SEQUENCE 530 AA; 57828 MW; 275363BF15440730 CRC64;

Query Match 32.1%; Score 660; DB 2; Length 530;
Best Local Similarity 39.2%; Pred. No. 3.6e-47;
Matches 145; Conservative 68; Mismatches 123; Indels 34; Gaps 9;

Qy 36 GGEARVHLVLSGWSGSSFGQLFNQHPDVFYLMPEAHVWTTLSQGSAAATLHMVRDL 95
Db 160 GGDKRQLVYVFTTWRSWGSSFFGELFNQNPVEFFLYEPVHVWQKLYPGDAVSLQGAARDM 219
Qy 96 VRSVFLCDMDVFAYLP---WRNLSDL---FQWAVSRALCSPACSAFPRGALS--SEAV 148
Db 220 LSALYRCDLSVFLQYSPAGSGGRNLTTLGFGAATNKVVCSSPLCPAYRKEVVGLVDDRV 279
Qy 149 CKPLCARQSFLLAREACRSYSHVVLKEVRFNQLVPLSDPALNLRIVHLVRDPRAYL 208
Db 280 CKK-CPQRLARFECECRKYRTLVIKGVRFVDVAVLAPLLKDPALDLKVHLVRDPRAVA 338
Qy 209 RSREQTAKALARDN-----GIVLGT--NGTWVEAD-PGLRVVREV 245
Db 339 SSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAAGHKLGAKGEGMGADYHALGAMEVI 398
Qy 246 CRSHVRIAEATLKPPPLRGRLVRFEDLAREPLAEIRALYAFGLSLTPQLEAWIHN 305
Db 399 CNSMAKTLQTA-LQPPDLQGHYLVRYEDLVGDPVKTLRRVYDFVGLLVSPMEQFALN 457
Qy 306 ITHGSGPGARREAFKTSRRNALNVSOAHRHALPFKIRRVQELCAGALQLLGYRVPVSED 365
Db 458 MTSGSGSSSK--PFVVSARNATQANANWRTALTFOQIKQVEEFCYQPMVILGYRVNSPE 515
Qy 366 EORNALDLV 375
Db 516 EVKDLSTLL 525

RESULT 14
Q9UED5 PRELIMINARY; PRT; 483 AA.
ID Q9UED5
AC Q9UED5;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
GN Name=GN6ST;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawa A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells."; J.
RL J. Biochem. 124:670-678(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=20498786; PubMed=11042394; DOI=10.1016/S0304-4165(00)00136-7;
RA Sakaguchi H., Kitagawa H., Sugahara K.;
RT "Functional expression and genomic structure of human N-
RT acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT sequence."; Biochim. Biophys. Acta 1523:269-276(2000).
RL Biochim. Biophys. Acta 1523:269-276(2000).
DR EMBL; AB014679; BAA34265.2; -
DR EMBL; AB021124; BAB16886.1; -
DR EMBL; AB021125; BAB16887.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;

Query Match 32.0%; Score 657.5; DB 2; Length 483;
Best Local Similarity 38.2%; Pred. No. 5.2e-47;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;

Qy 30 GPSSPAG-----GEARVHVLVLSWRSGSFVQGLFNQHPDVFYLMPEPAHVVMTTL 80
Db 98 GVAAPPNGTGTGGVGDKEQLVYVFTTWSGSGSFGELEFNQNPVEVFLYEPVHVHVKQL 157

Qy 81 SQGSAATLHMVRLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPAC 135
Db 158 YPGDAVSLQGAARDMLSALYRCDLSVQLYSPAGSGGNLTTLGIFGAATNKVVCSSPLC 217

Qy 136 SAFPRGAIS--SEAVCKPLCARQSFLLAREACRSYSHVVLKEVFFNQLVPLLSDPAL 193
Db 218 PAYRKEVVGVLVDVCKK-CPQRLARFEBCRKYRTLVKGVRFVDVAVLAPLRDPAL 276

Qy 194 NLRIVHLVDRPRAVLRSEQTAKALARDN-----GIVLGT--NGT 231
Db 277 DLKVIHLVDRPRAVASSRIRSRHGLRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGV 336

Qy 232 WVEAD-PGLRVVRECVSHVRIAATAATLKPFPFLGRVRLVRFFEDLAREPLAEIRALYAF 290
Db 337 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRYDF 395

Qy 291 TGLSLTPQLEAWIHNTHGSGPGARREAFKTSRNALNVSOAWRHAPLFAKIRRVQELCA 350
Db 396 VGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATQAANAWRTALTFTQIKQVEBFY 453

Qy 351 GALQLLGYRPVYSEDEQRLALDLV 375
Db 454 QPMVGLGYRVPVSEDEQRLALDLV 478

RESULT 15
QY4C5 PRELIMINARY; PRT; 530 AA.
ID QY4C5; Q9GZNS; Q9Y6F2;
AC QY4C5; Q9GZNS; Q9Y6F2; 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 27, Last annotation update)
DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST)
DE Carbohydrate sulfotransferase 2).
GN Name=GN6ST; Synonyms=CHST2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98391845; PubMed=9722682;
RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
RA Ozaki T., Nakagawa A., Kadamatsu K., Muramatsu T.,
RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
RT biosynthesis of 6-sulfo sialyl Lewis X: molecular cloning, chromosomal
RT mapping, and expression in various organs and tumor cells."; J.
RL J. Biochem. 124:670-678(1998).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein endothelium;
RX MEDLINE=99168906; PubMed=10049591; DOI=10.1006/geno.1998.5653;
RA Li X., Tedder T.F.;
RT "CHST1 and CHST2 sulfotransferases expressed by human vascular
RT endothelial cells: cDNA cloning, expression, and chromosomal
RT localization."; Genomics 55:345-347(1999).
RL Genomics 55:345-347(1999).
DR EMBL; AB014680; BAA34266.2; -
DR EMBL; AF083066; AAD20981.1; -
DR Genew; HGNC:1970; CHST2.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase..
DR Pfam; PF00685; Sulfotransfer_1; 1.
KW Transferase.
SQ SEQUENCE 530 AA; 57857 MW; AB2CA227B9D5651B CRC64;

Query Match 32.0%; Score 657.5; DB 2; Length 530;
Best Local Similarity 38.2%; Pred. No. 5.9e-47;
Matches 147; Conservative 70; Mismatches 125; Indels 43; Gaps 10;

Qy 30 GPSSPAG-----GEARVHVLVLSWRSGSFVQGLFNQHPDVFYLMPEPAHVVMTTL 80
Db 145 GVAAPPNGTGTGGVGDKEQLVYVFTTWSGSGSFGELEFNQNPVEVFLYEPVHVHVKQL 204

Qy 81 SQGSAATLHMVRLVRSVFLCDMDVFDAYLP---WRRNLSDL--FQWAVSRALCSPAC 135
Db 205 YPGDAVSLQGAARDMLSALYRCDLSVQLYSPAGSGGNLTTLGIFGAATNKVVCSSPLC 264

Qy 136 SAFPRGAIS--SEAVCKPLCARQSFLLAREACRSYSHVVLKEVFFNQLVPLLSDPAL 193
Db 265 PAYRKEVVGVLVDVCKK-CPQRLARFEBCRKYRTLVKGVRFVDVAVLAPLRDPAL 323

Qy 194 NLRIVHLVDRPRAVLRSEQTAKALARDN-----GIVLGT--NGT 231
Db 324 DLKVIHLVDRPRAVASSRIRSRHGLRESLQVRSRDPRAHRMPFLEAAGHKLGAKEGV 383

Qy 232 WVEAD-PGLRVVRECVSHVRIAATAATLKPFPFLGRVRLVRFFEDLAREPLAEIRALYAF 290
Db 384 GGPADYHALGAMVEICNSMAKTLQTA-LQPPDWLQGHVLYVRYEDLVGDPVKTLRRYDF 442

Qy 291 TGLSLTPQLEAWIHNTHGSGPGARREAFKTSRNALNVSOAWRHAPLFAKIRRVQELCA 350
Db 443 VGLLVSPMEQFALNMTSGSGSSK--PFVVSARNATQAANAWRTALTFTQIKQVEBFY 500

Qy 351 GALQLLGYRPVYSEDEQRLALDLV 375

Search completed: June 23, 2005, 08:49:01
Job time : 44.3472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:32:43 ; Search time 132.064 Seconds
(without alignments)
3578.727 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPRGGAPPIMALMFTGHL.....LIENICWTLMDRLGYPKFM D 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6532	100.0	1222	4	AAV72642 Human gly
2	6509	99.6	1222	6	ABU11849 Human sec
3	6465	99.0	1212	6	AAE33541 Human nov
4	5899.5	90.3	1207	4	AAV72643 Mouse gly
5	5899.5	90.3	1207	6	AAE33542 Human nov
6	4037	61.8	755	4	AAAB93735 Human pro
7	4037	61.8	755	8	ADP56673 Human pro
8	3147	48.2	596	4	AAV72641 Human gly
9	2597	39.8	480	4	AAAB94410 Human pro
10	1753.5	26.8	958	3	AAAY51120 Human SAR
11	1753.5	26.8	958	8	ADL83102 Human PRO
12	1331	20.4	314	4	AAAB93826 Human pro
13	898	13.7	625	8	ADR10287 Human pro
14	637	9.8	128	4	ADG27796 Human nov
15	568	8.7	125	8	ADJ12217 Human sec
16	568	8.7	126	2	AAAY14415 Human sec
17	431	6.6	456	8	ADP30411 Human sec
18	368	5.6	218	4	AAM24176 Human EST
19	348.5	5.3	474	3	AAAB34722 Human sec
20	147	2.3	410	7	ADN95522 Human BEC
21	147	2.3	411	2	AAW61100 Keratan s
22	147	2.3	411	5	AAE25356 Human cho
23	147	2.3	411	6	ABU03503 Angiogene
24	147	2.3	411	7	ADJ68589 Human hea
25	147	2.3	411	8	ADQ18725 Human sof

26	145	2.2	2504	2	AAW17871	AAW17871 Photorhab
27	145	2.2	2504	2	AAW56557	AAW56557 Toxin Tcb
28	145	2.2	2504	4	AAAB72610	AAAB72610 Photorhab
29	145	2.2	2504	5	ABG32654	ABG32654 P. lumine
30	145	2.2	2504	8	ADR21539	ADR21539 Photorhab
31	145	2.2	2505	4	AAAB72612	AAAB72612 Modified
32	135	2.1	1844	2	AAW18302	AAW18302 Photorhab
33	135	2.1	1844	2	AAW56558	AAW56558 Toxin Tcb
34	134	2.1	483	2	AAV31656	AAV31656 Mouse N-a
35	133	2.0	484	2	AAV31657	AAV31657 Human N-a
36	133	2.0	530	4	AAAB95367	AAAB95367 Human pro
37	133	2.0	530	8	ADQ18590	ADQ18590 Human sof
38	133	2.0	531	5	AAU69414	AAU69414 Lung smal
39	133	2.0	531	8	ADRI14283	ADRI14283 Human NF-
40	130	2.0	1456	2	AAAR24033	AAAR24033 Soluble m
41	128	2.0	183	4	ABAB68582	ABAB68582 Drosophil
42	127	1.9	1456	6	ABU79121	ABU79121 Angiostat
43	127	1.9	1456	7	ADPA43375	ADPA43375 Mannose r
44	127	1.9	1456	7	ADN39899	ADN39899 Cancer/an
45	127	1.9	1456	7	ADN95147	ADN95147 Human BEC

ALIGNMENTS

RESULT 1

AAV72642

ID AAV72642 standard; protein; 1222 AA.

XX AAV72642;

XX 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-6 (GST-6).

Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Domain 851..1223

XX /label= C-terminal_sulfotransferase_domain

XX WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02702, AAD02704.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 5B; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-6 (GST-6). GST is

CC a type 2 membrane protein useful for inhibiting a binding event between a
 CC selectin and a selectin ligand, which comprises contacting the selectin
 CC with a non-sulphated selectin ligand, GST and a small molecular agent
 CC that inhibits the sulphation activity of GST. GST is also useful in
 CC inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, perni-
 CC c dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation
 XX

SQ Sequence 1222 AA;

Query Match	100.0%;	Score 6532;	DB 4;	Length 1222;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1222;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	MPKGGAPPIMALMFTGHLFLALLMFAFSTPEESVSNYSENAVFTDDIDQFKTKQVDF	60
Db	1	MPKGGAPPIMALMFTGHLFLALLMFAFSTPEESVSNYSENAVFTDDIDQFKTKQVDF	60
Qy	61	RPNQKLKSMHPSLYFDAGEIQAMRQKSPASHLHLFRAIRSAVTVMLSNPYYLPPPKH	120
Db	61	RPNQKLKSMHPSLYFDAGEIQAMRQKSPASHLHLFRAIRSAVTVMLSNPYYLPPPKH	120
Qy	121	ADFAAKWNEIYGNLPPALYCLLCPEDKVAFFVLEYMWRMVGKDWLVENAPGDEVPI	180
Db	121	ADFAAKWNEIYGNLPPALYCLLCPEDKVAFFVLEYMWRMVGKDWLVENAPGDEVPI	180
Qy	181	GHSLTGFATAFDLYNLNDNHRKQYLEKIWIITEEMEYKSVRSWKGKQLLNHQATNMI	240
Db	181	GHSLTGFATAFDLYNLNDNHRKQYLEKIWIITEEMEYKSVRSWKGKQLLNHQATNMI	240
Qy	241	ALLTGALVTGVDGSKANIWKQAVVDMKTMFLNHHIVDGLDEGVAIGSYTAKSVTQY	300
Db	241	ALLTGALVTGVDGSKANIWKQAVVDMKTMFLNHHIVDGLDEGVAIGSYTAKSVTQY	300
Qy	301	VFLAQRFHNINLNDNWKQHFYIYATLLPGFQRTVGIADSNYNWFGPESQVFLDKP	360
Db	301	VFLAQRFHNINLNDNWKQHFYIYATLLPGFQRTVGIADSNYNWFGPESQVFLDKP	360
Qy	361	ILKXGAGNWLAAQIRKXRPKDGMPVSTAQRWSTLHTEYIWDYDQPTQPPADYGTAKTH	420
Db	361	ILKXGAGNWLAAQIRKXRPKDGMPVSTAQRWSTLHTEYIWDYDQPTQPPADYGTAKTH	420
Qy	421	TFPNWGVVTVGAGLPNTQTNTVTFSPKSGKLGRNAVYDIVHFQPSYWDGWRSPNGHEHP	480
Db	421	TFPNWGVVTVGAGLPNTQTNTVTFSPKSGKLGRNAVYDIVHFQPSYWDGWRSPNGHEHP	480
Qy	481	DQNSFTFAPNGQVVFSEALYGPKLSHLNVLFAPSPSSQCNKPWEGQGECAQMLKWTG	540
Db	481	DQNSFTFAPNGQVVFSEALYGPKLSHLNVLFAPSPSSQCNKPWEGQGECAQMLKWTG	540
Qy	541	EEVGDAAGEIITASQHGEMVVFVSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIERQ	600
Db	541	EEVGDAAGEIITASQHGEMVVFVSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIERQ	600
Qy	601	EDSPINSVSAPFNLDIDFKYIIPKFNRYNGAMVDWAHYOMFWDHGHGNSPMASIOE	660
Db	601	EDSPINSVSAPFNLDIDFKYIIPKFNRYNGAMVDWAHYOMFWDHGHGNSPMASIOE	660
Qy	661	AEQAAEFKKRWTFQVNVVTFQMESTITRIAVVFYGPYINVSSCRFIDSSNPGIQLSNVNN	720
Db	661	AEQAAEFKKRWTFQVNVVTFQMESTITRIAVVFYGPYINVSSCRFIDSSNPGIQLSNVNN	720
Qy	721	TEHWVSIVTDYHNLKTRFNLYLGGFGFASVADQGIQITRFGLGTQAIKVPVRHDIRIIPFGF	780
Db	721	TEHWVSIVTDYHNLKTRFNLYLGGFGFASVADQGIQITRFGLGTQAIKVPVRHDIRIIPFGF	780

Qy	781	KFNIAVGLILCISLVILTFQWRFYLSFRKLMRWILILVIALWFIELLDVWSTCSQPICAK	840
Db	781	KFNIAVGLILCISLVILTFQWRFYLSFRKLMRWILILVIALWFIELLDVWSTCSQPICAK	840
Qy	841	WTRTEAGSKLSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVPTAYIDI	900
Db	841	WTRTEAGSKLSLSEGHMDLPDVVITSLPGSGAEILKQLFFNSDDFLYIRVPTAYIDI	900
Qy	901	PETELEIDSFVDACSWKVSVDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHHPNRKLAQY	960
Db	901	PETELEIDSFVDACSWKVSVDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHHPNRKLAQY	960
Qy	961	FAMNKDKRKKRRESLPQRSQMGKAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGSWT	1020
Db	961	FAMNKDKRKKRRESLPQRSQMGKAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGSWT	1020
Qy	1021	LKLHFFQEVGLGASMRALYIVRDPRAWIYSLMYNSKPSLYSLKNVPEHLAKLFKIEGGKKG	1080
Db	1021	LKLHFFQEVGLGASMRALYIVRDPRAWIYSLMYNSKPSLYSLKNVPEHLAKLFKIEGGKKG	1080
Qy	1081	CNLSGSAFAYEPEPRKELSKSKSNVSLLSHLWLANTAAALRINTDLLPTSYQLVKPFEDI	1140
Db	1081	CNLSGSAFAYEPEPRKELSKSKSNVSLLSHLWLANTAAALRINTDLLPTSYQLVKPFEDI	1140
Qy	1141	VHFPOKTTTIRIFAFILGIPLSASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDE	1200
Db	1141	VHFPOKTTTIRIFAFILGIPLSASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDE	1200
Qy	1201	IKLIENICWTLMDRILGYPKFMD 1222	
Db	1201	IKLIENICWTLMDRILGYPKFMD 1222	

RESULT 2	ABU11849
ID	ABU11849 standard; protein; 1222 AA.
XX	ABU11849;
AC	ABU11849;
XX	
DT	12-FEB-2003 (first entry)
XX	
DE	Human secreted protein SECP-4, INCYTE 3441255CD1.
KW	Human; SECP; secreted protein; micro-array; liver disease; hepatitis; cirrhosis; cell proliferative disease; cancer; atherosclerosis;
KW	neurological disorder; epilepsy; Huntington's disease; stroke;
KW	cardiovascular disorder; hypertension; angina pectoris; allergy;
KW	myocardial infarction; immune disorder; inflammatory disorder; AIDS;
KW	hypothyroidism; acquired immunodeficiency syndrome; Cushing's syndrome; developmental disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200286069-A2.
XX	
PD	31-OCT-2002.
XX	
PF	19-APR-2002; 2002WO-US012464.
XX	
PR	20-APR-2001; 2001US-0285207P.
PR	27-APR-2001; 2001US-0287114P.
PR	03-MAY-2001; 2001US-0288640P.
PR	11-MAY-2001; 2001US-0290516P.
PR	18-MAY-2001; 2001US-0292184P.
PR	21-DEC-2001; 2001US-0343553P.
PR	13-FEB-2002; 2002US-0357002P.
PR	20-FEB-2002; 2002US-0358279P.
PR	19-MAR-2002; 2002US-0366041P.
XX	
PA	(INCY-) INCYTE GENOMICS INC.
XX	
PI	Klammer AA, Hafalia AJA, Duggan BM, Warren BA, Emerling BM;

PI Tribouley CM, Arvizu CS, Honchell CD, Nguyen DB, Kallick DA, Yue H;
 PI Au-Young JK, Ramkumar J, Li JX, Thangavelu K, Gietzen KJ, Ding L;
 PI Baughn MR, Yao MG, Malia NK, Mason PM, Lal PG, Graul RC, Reddy R;
 PI Becha SD, Sapperstein SK, Richardson TW, Tran UK, Elliott VS;
 PI Tang YT, Azimzai Y, Yan L, Xu Y;
 XX WPI; 2003-093118/08.
 DR N-PSDB; ABX49951.
 XX
 PT New human secreted proteins (SECP) useful for diagnosing, treating and
 PT preventing diseases or conditions associated with the aberrant SECP
 PT expression e.g. cancer, AIDS, atherosclerosis, epilepsy, allergies,
 PT hepatitis, cirrhosis.
 XX
 PS Claim 1; Page 137-140; 192pp; English.
 XX
 CC The invention relates to an isolated polypeptide comprising any of 30
 CC secreted human proteins (SECP1-SECP30) appearing as ABU1846-ABU1875, a
 CC naturally occurring amino acid sequence at least 90-98 % identical to the
 CC sequences, or a biologically active or immunogenic fragment of the
 CC polypeptide. Also included are an isolated polynucleotide encoding SECP,
 CC (including a polynucleotide sequence at least 90-98 % identical to the
 CC sequences, their complements, RNA equivalents or fragments comprising at
 CC least 60 contiguous nucleotides) a recombinant polynucleotide comprising
 CC a promoter sequence operably linked to the SECP polynucleotide, a cell
 CC comprising the recombinant polynucleotide, a transgenic organism
 CC comprising the recombinant polynucleotide, an anti-SECP antibody,
 CC screening for ant/agonists of SECP, generating an expression profile of a
 CC sample containing the polynucleotides and an array comprising different
 CC nucleotide molecules affixed at distinct physical locations on a solid
 CC substrate, where at least one nucleotide molecule comprises a first
 CC oligonucleotide or polynucleotide sequence specifically hybridisable with
 CC at least 30 contiguous nucleotides of the target polynucleotide. The
 CC polypeptides and polynucleotides are useful in diagnosing, treating and
 CC preventing diseases or conditions associated with the decreased
 CC expression or overexpression of SECP, such as liver (e.g. hepatitis,
 CC cirrhosis), cell proliferative (e.g. cancer, atherosclerosis),
 CC neurological (e.g. epilepsy, Huntington's disease, stroke),
 CC cardiovascular (e.g. hypertension, angina pectoris, myocardial
 CC infarction), immune/inflammatory (e.g. acquired immunodeficiency syndrome
 CC (AIDS), allergies) and developmental (e.g. Hypothyroidism, Cushing's
 CC syndrome) disorders (many other diseases and conditions are given in the
 CC specification). These are also useful in assessing the effects of
 CC exogenous compounds on the expression of nucleic acid and amino acid
 CC sequences of SECP. The SECP or its fragments are useful in screening
 CC compounds for effectiveness as agonist or antagonist of the polypeptides,
 CC or in altering the expression of the target polynucleotide and compounds
 CC that specifically bind to or modulate the activity of the polypeptide.
 CC The micro-array is useful in monitoring or measuring protein-protein
 CC interactions, drug-target interactions, and gene expression profiles. The
 CC present sequence is a SECP protein of the invention
 XX
 SQ Sequence 1222 AA;

Query Match 99.6%; Score 6509; DB 6; Length 1222;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MPKGGAPPWIMALMFTGHLFLALLMFAFSTFEESVSNYSMAVFTDDIDQFKTKQVDF 60
 DB 1 MPKGGAPPWIMALMFTGHLFLALLMFAFSTFEESVSNYSMAVFTDDIDQFKTKQVDF 60
 QY 61 RPNOKLKSMHPSLYFDAGIOAMRQKSRASHLHLFRAISRATVMSLNPTYYLPPPKH 120
 DB 61 RPNOKLKSMHPSLYFDAGIOAMRQKSRASHLHLFRAISRATVMSLNPTYYLPPPKH 120
 QY 121 ADFAAKWNEIYGNLPLPPLALCYLLCPEDKVAFEFVLEYMDRMVGYKDWLVENAPGDEVPI 180
 DB 121 ADFAAKWNEIYGNLPLPPLALCYLLCPEDKVAFEFVLEYMDRMVGYKDWLVENAPGDEVPI 180
 QY 181 GHSLTGFATAPDFLYNLNDHRRQKYLEKIWIITEEMEYYSKVRSGWQQLLHNHQAATMI 240
 DB 181 VHSLTGFATAPDFLYNLNDHRRQKYLEKIWIITEEMEYYSKVRSGWQQLLHNHQAATMI 240

QY 241 ALLTGALVTGVDKSGKANIKQAVDVMKTMFLNHHIVDGLSDEBVGAYSYTAKSVTQY 300
 DB 241 ALLTGALVTGVDKSGKANIKQAVDVMKTMFLNHHIVDGLSDEBVGAYSYTAKSVTQY 300
 QY 301 VFLAQHFNNINNDNNWLKMHFWFYATLLPGFQRTVGIADSNYNWYFPGESQVFLDKP 360
 DB 301 VFLAQHFNNINNDNNWLKMHFWFYATLLPGFQRTVGIADSNYNWYFPGESQVFLDKP 360
 QY 361 ILKNGAGNMLAQOIRKHRPKDGMVSTAQRWSTLHTEYIWDYDQTPADYGTAKIH 420
 DB 361 ILKNGAGNMLAQOIRKHRPKDGMVSTAQRWSTLHTEYIWDYDQTPADYGTAKIH 420
 QY 421 TFPNWGVVITYGAGLPNTQTNTFVSFKSGKLGRAVDYIVHFQPYSWIDGWSRSPGHEHP 480
 DB 421 TFPNWGVVITYGAGLPNTQTNTFVSFKSGKLGRAVDYIVHFQPYSWIDGWSRSPGHEHP 480
 QY 481 DQNSFTFAPNGQVVFSEALYGPKLHLNNVLFVAPSPSSQCNKPWEGQLGECQWLKWTG 540
 DB 481 DQNSFTFAPNGQVVFSEALYGPKLHLNNVLFVAPSPSSQCNKPWEGQLGECQWLKWTG 540
 QY 541 EEVGDAAGEITITASOHEGMVFSGEAVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 DB 541 EEVGDAAGEITITASOHEGMVFSGEAVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 QY 601 EDSPINSVSFAFFHNLDIDFKYIYPKFMNRYNGAMDMVDADAHYKMFDFDHGNSPMASIQE 660
 DB 601 EDSPINSVSFAFFHNLDIDFKYIYPKFMNRYNGAMDMVDADAHYKMFDFDHGNSPMASIQE 660
 QY 661 AEQAASFKKWTFQVNVTFQMESTITRIAYVFGPYINVSCTPIDSNSPGLQISLVNWN 720
 DB 661 AEQAASFKKWTFQVNVTFQMESTITRIAYVFGPYINVSCTPIDSNSPGLQISLVNWN 720
 QY 721 TEHVSVITVDYHNLKTRFNVLGGGFASVADQOITRFGLGTAIVKPVHRDRIIFPFG 780
 DB 721 TEHVSVITVDYHNLKTRFNVLGGGFASVADQOITRFGLGTAIVKPVHRDRIIFPFG 780
 QY 781 KFNIAVGLIICISLVILTFCWRFLYSPKLMRWLILVIALWFIELDDVMSTCSQPICAK 840
 DB 781 KFNIAVGLIICISLVILTFCWRFLYSPKLMRWLILVIALWFIELDDVMSTCSQPICAK 840
 QY 841 WTRTEAGSKKSLSSSEGHMDLPDVTITSLPGSGAEILKOLFNSDFLIYRVPYATIDI 900
 DB 841 WTRTEAGSKKSLSSSEGHMDLPDVTITSLPGSGAEILKOLFNSDFLIYRVPYATIDI 900
 QY 901 PETELEIDSFVDACEWKSVDIRSGHFRLLRGMQSLVQDTKLHLQNLHLEPNRGLAQY 960
 DB 901 PETELEIDSFVDACEWKSVDIRSGHFRLLRGMQSLVQDTKLHLQNLHLEPNRGLAQY 960
 QY 961 FANMKDKKRFKRESLPEQORSQMKGAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGWT 1020
 DB 961 FANMKDKKRFKRESLPEQORSQMKGAFDRDAEYIRALRRHLVYYPSPARPVLSLSSGWT 1020
 QY 1021 LKLFHFQEVILGASNRALYIVRDPRAWIYMLYNSKPSLSYLNKVPHEHLAKLFKEGGK 1080
 DB 1021 LKLFHFQEVILGASNRALYIVRDPRAWIYMLYNSKPSLSYLNKVPHEHLAKLFKEGGK 1080
 QY 1081 CNLNSGYAFEPYELRKELSKSNVSLLSHLWLANTAARINTDLPSTSYQLVKREDI 1140
 DB 1081 CNLNSGYAFEPYELRKELSKSNVSLLSHLWLANTAARINTDLPSTSYQLVKREDI 1140
 QY 1141 VHPQKTERIFAPLGIPLSPASLNQILFATSTNLVFLPYVEGEISPTNTNVWKNLPRDE 1200
 DB 1141 VHPQKTERIFAPLGIPLSPASLNQILFATSTNLVFLPYVEGEISPTNTNVWKNLPRDE 1200
 QY 1201 IKLIENICWTLMDRLGYPKFMD 1222
 DB 1201 IKLIENICWTLMDRLGYPKFMD 1222
 RESULT 3
 AAE33541
 ID AAE33541 standard; protein; 1212 AA.

polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; aneamia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection.

Mus sp.
W0200106015-A1.
25-JAN-2001.
19-JUL-2000; 2000MO-US019741.
20-JUL-1999; 99US-0144694P.
13-JUN-2000; 2000US-00593828.
(REGC) UNIV CALIFORNIA.
Rosen SD, Lee JK, Hemmerich S;
WPI; 2001-138471/14.
N-PSDB; AAD02705, AAD02706.
New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications.
Example 2; Fig 6B; 128pp; English.
The present sequence is mouse glycosyl sulfotransferase-6 (GST-6). GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adenitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation

Query Match 90.3%; Score 5899.5; DB 4; Length 1207;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

QY 11 MALMFTGHLALFALLMPAFSTFEESVNSYSEWAVFTDDIDQFTKQVQDFRPNQKLKSM 70
DB 1 MAFNTEHLLFLLTLMSCSFSTCEESVNSYSEWAVFTDDIQWLKSKQIDFKLNR----- 55
QY 71 LHPSLYFDAGEIQAMROKQRASHLHLFRATRSVTVMLSNPTYYLPPLPPKQADPAKKNWEI 130
DB 56 LHPNLVFDAGDIQTLKQKSRSTSHLHIFRAIKSAVTIMLSNPSSYYLPPLPPKGAFAAKWNEI 115
QY 131 YGNLPLALYCLCPEDKVAFEVLEVMOMVGYKDWLVENAPGDEVPIGHSITGFATA 190
DB 116 YGNLPLALYCLCPEDKVAFEVLEVMOMVGYKDWLVENAPGDEVPIGHSITGFATA 175
QY 191 FDFLYNLLDNHRRQKYLEKIWIITEEMYEYSKVRSGKQLLHNNHQATNMIALLTGALVTG 250
DB 176 FDFLYNLLGNQRQKYLEKIWIITEEMYEYSKVRSGKQLLHNNHQATNMIALLTGALVTG 235
QY 251 VDKSGKANIWKQAVVDVNEKTMFLNHNHVDGSLDEGVAYSYTAKSVTQYVFLAQRHNI 310
DB 236 VDKSGKANIWKQAVVDVNEKTMFLNHNHVDGSLDEGVAYSYTAKSVTQYVFLAQRHNI 295

QY 311 NNLDNNWLKQHFYFATYATLLPGFORTVGADSNYNWYFPGESQVFLDFKILKNGAGNWL 370
DB 296 NNFDNNWLKQHFYFATYATLLPGVQRTVGADSNYNWYFPGESQVFLDFKILKNGAGNWL 355
QY 371 AQOIRKRPKQGMVPSSTAQRSTLHTEYIWDVDPQLTPOPPADYGTAKLHTEFNNWGVVY 430
DB 356 AQOIRKRPKQGMVPSSTAQRSTLHTEYIWDVDPQLTPOPPADYGTAKLHTEFNNWGVVY 415
QY 431 GAGLPNTQNTTFVSFKSGKLGGRVAVYDI VHFQPSYMWIDGWRSPNPGHEHPDQNSFTFAPN 490
DB 416 GGGLPNTQNTTFVSFKSGKLGGRVAVYDI VHFQPSYMWIDGWRSPNPGHEHPDQNSFTFAPN 475
QY 491 GQVFSEALYGPKLHNLNVLVFPSPSSQCNKPWEGQLGCECAQWLKWTGEYVGDAGEI 550
DB 476 GQVFSEALYGPKLHNLNVLVFPSPSSQCNKPWEGQLGCECAQWLKWTGEYVGDAGEI 535
QY 551 ITASQHGEMVYSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIEREDSPINSVSA 610
DB 536 ITAAQHGRMFVSGEAVSAYSSAMRLKSVYRALLLNSQTLVVDHIEREDSPINSVSA 595
QY 611 FPHNLIDIFKYIPYKPMNRYNGAMDMVDAHYKMFWDHHCNSPMASIQEAEQAAEFKKR 670
DB 596 FPHNLIDIFKYIPYKPMNRYNGAMDMVDAHYKMFWDHHCNSPMASIQEAEQAAEFKKR 655
QY 671 WTOPVNVTFOMESTITRIAYVYFYPYINVSSCRFDISSNPGLOISLNNVNTHEHVSIIVTD 730
DB 656 WTOPVNVTFOMESTITRIAYVYFYPYINVSSCRFDISSNPGLOISLNNVNTHEHVSIIVTD 715
QY 731 YHNLKRFNVLGFGFASVADOGITRFGLGTOAIYKVPVRHRIIPFPFKENIAGVLIL 790
DB 716 YQNLKSRFVYLGFGFASVANOGQITRFGLGTOEIVNVPVRHDKVNPFPFKFNIAVGFI 775
QY 791 CISLVLTQWRFLYLFKRLMRWILILVIALWFIELLDVNMSTCSQPICAKWTREABSGK 850
DB 776 CISLVLTQWRFLYLFKRLMRWILILVIALWFIELLDVNMSTCTQPICAKWTREABSGK 835
QY 851 KSLSSGHHMDLPDVVITSLPGSGAILKQLPFNSSDFLYRVPVTAVIDIPETELEDSP 910
DB 836 KWLSEGHVLDLPVITSLPGSGAILKQLPFNSSDFLYRVPVTAVIDIPETELEDSP 895
QY 911 VDACEWKSVDIRSGHFLRLGWLQSLVODTKLHLNHLHHPNKGKLAQYFANNKDKKRR 970
DB 896 VDACEWKSVDIRSGHFLRLGWLQSLVODTKLHLNHLHHPNKGKLAQYFANNKDKKRR 955
QY 971 FKRESLPEORSOMKGAFFORDAEYIRALRHLVYPSARVLSLSSGSWTLKLFHFOEVL 1030
DB 956 LKRRESLQDQSRKIGPFDDAEYIRALRHLVYPSARVLSLSSGSWTLKLFHFOEVL 1015
QY 1031 GASMRALYIVRDPRAWIYSMLYNSKPSLSYSLKNVPEHLAKLFIKGGKGCNLSGYAFE 1090
DB 1016 GTSMRALYIVRDPRAWIYSMLYNSKPSLSYSLKNVPEHLAKLFIKGGKGCNLSGYAFE 1075
QY 1091 YEPRLKELSKSNVSLSHLWLANATAALRINTDLPSTSYOLVKPEDIVHPPOKTER 1150
DB 1076 YESLKELEISQSNALSLSHLWLANATAALRINTDLPSTSYOLVKPEDIVHPPOKTER 1135
QY 1151 IFAPLGIPLSPASINOLIFATSTNLFLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
DB 1136 IFAPLGIPLSPASINOLIFATSTNLFLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
QY 1211 LMDRLGYPKFMD 1222
DB 1196 LMDHLGYPKFMD 1207
RESULT 5
AAE33542
ID AAE33542 standard; protein; 1207 AA.
XX AAE33542;
AC AAE33542;
XX
DT 16-APR-2003 (first entry)
XX

DE Human novel CpG-associated gene 1 (NCAG1) encoded protein #2.
KW Human; novel CpG-associated gene 1; bipolar disorder; neuroprotective;
KW NCAG1; mood disorder; chromosome 18.
XX
OS Homo sapiens.
XX WO2002101044-A2.
XX
XX 19-DEC-2002.
XX
XX 06-JUN-2002; 2002WO-EP006316.
XX
XX 11-JUN-2001; 2001EP-00202214.
XX (JANC ; JANSSEN PHARM NV.
XX Del-Favero JPL, Van Broeckhoven C;
XX WPI; 2003-148807/14.
XX N-PSDB; AAD50033.
XX
XX New brain expressed genes (designated novel CpG-associated Gene 1
PT (NCAG1)) and its encoded protein, useful as diagnostic markers for
PT bipolar or mood disorders, and as targets for developing drugs for the
PT treatment these disorders.
XX
XX Claim 16; Col 48-51; 26pp; English.
XX
XX The invention relates to novel CpG-associated gene 1 (NCAG1) brain-
CC expressed gene and its encoded protein. The NCAG1 nucleic acid or its
CC encoded protein is useful as a diagnostic marker for bipolar disorder
CC such as mood disorders. They are also useful as targets for developing
CC drugs, as well as for target validation, for the treatment of bipolar
CC disorders. The present sequence is human NCAG1 protein. NCAG1 gene is
CC located at chromosome 18
XX
SQ Sequence 1207 AA;

Query Match 90.3%; Score 5899.5; DB 6; Length 1207;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

QY 11 MALMFTGHLFLALLMFAFTFEESVNTSEWAVFTDDIDQFTQKQVQDFRPNQKLKSM 70
DB 1 MAFMFTHEHLFLTLWCSTCEESVNTSEWAVFTDDIQWLKSKQIQDFKLNRR----- 55

QY 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPPHADFAAKWNEI 130
DB 56 LHPNLVFDAGDIQTLKQKSRSTSHLHIFRAIKSAVTIMLSNPSYLLPPPKHAEFAAKWNEI 115

QY 131 YGNLPLALYCLLPEDKVAFFEVLYMDRMVGYKDWLENAPGDGDEVPDIGHSLTGFA 190
DB 116 YGNLPLALYCLLPEDKVAFFEVLYMDRMVGYKDWLENAPGDGDEVPVGHSLTGFA 175

QY 191 FDELYNLLDNHRQKYLEKIWTIEMEYEVKRSWKGKOLLHNHQAATNMIALLTGALVTG 250
DB 176 FDELYNLLGNQRQKYLEKIWTIEMEYEVKRSWKGKOLLHNHQAATNMIALLIGALVTG 235

QY 251 VDKGSKANIWKQAVDVMDEKTMFLNHI VDGSLDEGVAYGYSYAKSVTYQVFLAQRHFN 310
DB 236 VDKGSKANIWKQAVDVMDEKTMFLNHI VDGSLDEGVAYGYSYAKSVTYQVFLAQRHFN 295

QY 311 NNLDNNLWKHFWFYATLLPGFQRTVGADSNYNWYFPGESQLVFLDKFLKNGAGNWL 370
DB 296 NNFDNNLWKHFWFYATLLPGYQRTVGADSNYNWYFPGESQLVFLDKFILLQNGAGNWL 355

QY 371 AQQIRKRPKDGPMVSTAQRTWTLHTEYIWDYDPTLPQADYGTAKHTPPNMGVVTY 430
DB 356 AQQIRKRPKDGPMVSTAQRTWTLHTEYIWDYDPTLPQADYGTAKHTPPNMGVVTY 415

QY 431 GAGLPNTQNTFVSFKSGKLGRAVYDIVHFQYPSWIDGMRSPNPGCHEHPDQNSFTFAPN 490
DB

DB 416 GGGLPNTQNTFVSFKSGKLGRAVYDIVHFQYPSWIDGMRSPNPGCHEHPDQNSFTFAPN 475
QY 491 GQVFVSEALYGPKLSHLANNLVAFAPSPSSQCNKPWEGOLGECAGQWLKWTGBEVGDAAGRI 550
DB 476 GQVFVSEALYGPKLSHLANNLVAFAPSPSSQCNKPWEGQLGECAGQWLKWTGBEVGDAAGRI 535
QY 551 ITASQHGEMVFSGEAVSAYSAMRLKSVYRALLLNSQTLVVDHIHQEDSPINSVSA 610
DB 536 ITAAQHGDRMFVSGEAVSAYSAMRLKSVYRALLLNSQTLVVDHIHQEDSPINSVSA 595
QY 611 PFHNLIDIDFKYIPYKPMRYNGAMVDVDAHYKMFWDHGHGNSPMASIOEABQAAEFKKR 670
DB 596 PFHNLIDIDFKYIPYKPMRYNGAMVDVDAHYKMFWDHGHGNSPMASIOEABQAAEFKKR 655
QY 671 WTQFVNVTQFMESTITRIAYVYFGYINVSRCRFDSSNPGLOISLVNNTNTHVVSIVTD 730
DB 656 WTQFVNVTQFMESTITRIAYVYFGYINVSRCRFDSSNPGLOISLVNNTNTHVVSIVTD 715
QY 731 YHNLKTRFNLYGFGGFASVADQGITRFGGLGTQAIKVPVRHDIIFPFGKFNIAVGLIL 790
DB 716 YQNLKSRFSYLGFGGFASVADQGITRFGGLGTQAIKVPVRHDIIFPFGKFNIAVGLIL 775
QY 791 CISLVILTFQWRFYLSFRKLNRWILVIALWFIELLDVWSTCSQPICAKWTTRTBAESK 850
DB 776 CISLVILTFQWRFYLSFRKLNRWILVIALWFIELLDVWSTCSQPICAKWTTRTBAESK 835
QY 851 KSLSEGHMMDLPDVVITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSF 910
DB 836 KVMISEGHVVDLPNVITSLPGSGAEILKOLFNSDFLYIRVPTAYIDIPETELEIDSF 895
QY 911 VDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFNPKGLAQYFAMNKKRKK 970
DB 896 VDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHFNPKGLAQYFAMNKKRKK 955
QY 971 FKRRSLPEQSSOMKAGFDROAEYIRALRHLVYVPSARPVLSLSSGWTILKHFQEV 1030
DB 956 LKRRESLQDRSRIKPGPDROAEYIRALRHLVYVPSARPVLSLSSGWTILKHFQEV 1015
QY 1031 GASMEALYIVRDPRAWIYSLMYNKPISLYLKNVPEHLAKLFKEGGKGNLNSGYAFE 1090
DB 1016 GTSMEALYIVRDPRAWIYSLMYNKPISLYLKNVPEHLAKLFKEGGKGNLNSGYAFE 1075
QY 1091 YEPLRKELSKSNASVLSLHLWLAATAALRINTDLPSTSYQLVKPEDI VHFPOKTTTER 1150
DB 1076 YESLKELEISQSNAILSLHLWLAATAALRINTDLPSTSYQLVKPEDI VHFPOKTTTER 1135
QY 1151 IPAFILGIPLSPASLNOILFATSTNLFYLPYGEISPTNWNWKNLPRDEIKLIENICWT 1210
DB 1136 IPAFILGIPLSPASLNOILFATSTNLFYLPYGEISPTNWNWKNLPRDEIKLIENICWT 1195
QY 1211 LMDRLGYPKFMD 1222
DB 1196 LMDHLGYPKFMD 1207

RESULT 6
AAB93735
ID AAB93735 standard; protein; 755 AA.
XX
AC AAB93735;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13382.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX

28-JUL-2000; 2000EP-00116126.
29-JUL-1999; 99JP-00248036.
27-AUG-1999; 99JP-00300253.
11-JAN-2000; 2000JP-00118776.
02-MAY-2000; 2000JP-00183767.
09-JUN-2000; 2000JP-00241899.
(HELI-) HELIX RES INST.
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.
Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
Claim 8; SEQ ID NO 13382; 2537pp + Sequence Listing; English.
The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
SQ Sequence 755 AA;
Query Match 61.8%; Score 4037; DB 4; Length 755;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 750; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 217 MYEYKVRWQKQLLNHNQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLN 276
Db 1 MYEYKVRWQKQLLNHNQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLS 60
Qy 277 HIVDGLDEGVAYGYSYAKSTQYVFLAQRHFNINLNNLWKKHFFYYATLPGFQRT 336
Db 61 HIVDGLDEGVAYGYSYAKSTQYVFLAQRHFNINLNNLWKKHFFYYATLPGFQRT 120
Qy 337 VGIADSNYNWFYGPESQLVFLDKFLKNGAGNLAQRIKHPKXGDMVPSQAQWSTLH 396
Db 121 VGIADSNYNWFYGPESQLVFLDKFLKNGAGNLAQRIKHPKXGDMVPSQAQWSTLH 180
Qy 397 TEYIWDPOLTPQPPADYGTAKIHFTFPNMGVYVYAGLPNTQTNTVFSFKSGKLGGRVY 456
Db 181 TEYIWDPOLTPQPPADYGTAKIHFTFPNMGVYVYAGLPNTQTNTVFSFKSGKLGGRVY 240
Qy 457 DIVHFPQSWIDGWRSNFPGHEHPDQNSFTAPNGQVFVSALYGPKLSHLNNLVFAPS 516
Db 241 DIVHFPQSWIDGWRSNFPGHEHPDQNSFTAPNGQVFVSALYGPKLSHLNNLVFAPS 300
Qy 517 PSSQCNKPEWQGLGCAQWLKWTGEEVDAAGEIITASQHGEMVFGSGEAVSAYSSAMRL 576

Db 301 PSQCNKPEWQGLGCAQWLKWTGEEVDAAGEIITASQHGEMVFGSGEAVSAYSSAMRL 360
Qy 577 KSVYRALLNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPYKPMNRYNGAMMD 636
Db 361 KSVYRALLNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPYKPMNRYNGAMMD 420
Qy 637 VMDAHYKMFWDHNGNSPMASIOEAQAAEFKRGWTQFVNVTFQMESTIIRIAYVFYGPY 696
Db 421 VMDAHYKMFWDHNGNSPMASIOEAQAAEFKRGWTQFVNVTFQMESTIIRIAYVFYGPY 480
Qy 697 INVSSCRFDSSNPGLOISLVNNTNTHVVSIVTDYHNLKTRNVLGFGGSPASVADQOQIT 756
Db 481 INVSSCRFDSSNPGLOISLVNNTNTHVVSIVTDYHNLKTRNVLGFGGSPASVADQOQIT 540
Qy 757 RFLGLGTQAIKVRHRIIPFGFKFNIAVGLICISLVILTQWRFYLSFRKLMRWILI 816
Db 541 RFLGLGTQAIKVRHRIIPFGFKFNIAVGLICISLVILTQWRFYLSFRKLMRWILI 600
Qy 817 LVIALWFIELLDVWSTCSQPICAKWRTRTEAEGSKLSSEGHHMDLPDVVITSLPGSGAE 876
Db 601 LVIALWFIELLDVWSTCSQPICAKWRTRTEAEGSKLSSEGHHMDLPDVVITSLPGSGAE 660
Qy 877 ILKQLFNSDFLYIRVPTAYIIDIPETELEDISFVDACEWKSVDIRSGHFLRGLQSL 936
Db 661 ILKQPFNSDFLYIRVPTAYIIDIPETELEDISFVDACEWKSVDIRSGHFLRGLQSL 720
Qy 937 VQDTKLHLQNIHLHPENRGKLAQYFAMNKDKKXK 970
Db 721 VQDTKLHLQNIHLHPENRGKLAQYFAMNKDKKXK 754
RESULT 7
ADP56673
ID ADP56673 standard; protein; 755 AA.
XX
AC ADP56673;
XX
DT 12-AUG-2004 (first entry)
XX
DE Human protein which is a chondroitin 6 sulphotransferase 3 homologue.
XX
KW cancer detection; large intestine; oesophagus; stomach; lungs; pancreas;
KW liver; kidney; colon; human; chondroitin 6 sulphotransferase 3; CH6T3.
XX
OS Homo sapiens.
XX
PN JP2004147505-A.
XX
PD 27-MAY-2004.
XX
PF 28-OCT-2002; 2002JP-003112927.
XX
PR 28-OCT-2002; 2002JP-003112927.
XX
PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.
PA (FURE) FUJIREBIO KK.
PA (SEGG) SEIKAGAKU KOGYO CO LTD.
XX
XX WPI; 2004-434537/41.
DR N-PSDB; ADP56672.
XX
XX Novel nucleic acid such as DNA, useful as probe for identifying or
PT detecting cancerous tissue of esophagus, stomach, lungs, pancreas, liver,
PT kidney or colon, preferably large intestine tissue.
XX
PS Example 1; SEQ ID NO 2; 24pp; Japanese.
XX
CC The invention relates to a novel nucleic acid having 40-1000 base pairs
CC and comprising a sequence which is complementary to 41-2308 nucleotides
CC of a fully defined sequence of 3288 base pairs as given in the
CC specification. The polynucleotide of the invention may be useful for
CC rapid and highly reliable detection of cancerous tissue derived from the
CC large intestine, oesophagus, stomach, lungs, pancreas, liver, kidney or

CC colon. The current sequence is that of the human protein of the invention
CC which is a chondroitin 6 sulphotransferase 3 (Ch6T3) homologue.

XX
SQ Sequence 755 AA;

Query Match 61.8%; Score 4037; DB 8; Length 755;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 750; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 217 MYEYSKVRSGKQLLNHQAETNIALTGTGVDKSGKANIWKQAVDVMKETHFLN 276
DB 1 MYEYSKVRSGKQLLNHQAETNIALTGTGVDKSGKANIWKQAVDVMKETHFLN 60
QY 277 HIVDGSDEGVAGSYTAKSVTVQVFLAQRHENINLNNLWKLHFWFYATLLPGFORT 336
DB 61 HIVDGSDEGVAGSYTAKSVTVQVFLAQRHENINLNNLWKLHFWFYATLLPGFORT 120
QY 337 VGIADSNYNFYGPESQLVFLDKFILKNGAGNLAQQIRKRPKDGMPVPSTQAQWSTLH 396
DB 121 VGIADSNYNFYGPESQLVFLDKFILKNGAGNLAQQIRKRPKDGMPVPSTQAQWSTLH 180
QY 397 TEIWDYDQPTPPADYGTAKHTPPNMGVYTYGAGLPNTQNTTFVSKGKLGGRVY 456
DB 181 TEIWDYDQPTPPADYGTAKHTPPNMGVYTYGAGLPNTQNTTFVSKGKLGGRVY 240
QY 457 DIVHFQYSNIDGWRSPNPGHEHPDQNSFTFAPNGOVFVSEALYGPKLHLNNVLFPAS 516
DB 241 DIVHFQYSNIDGWRSPNPGHEHPDQNSFTFAPNGOVFVSEALYGPKLHLNNVLFPAS 300
QY 517 PSSQCNKPEWGQCECAQWLKWTGEEVDAAGBIITASQHGEMFVSGEAVSAYSSAMRL 576
DB 301 PSSQCNKPEWGQCECAQWLKWTGEEVDAAGBIITASQHGEMFVSGEAVSAYSSAMRL 360
QY 577 KSVYRALLLNQTLVVDHIERQEDSPINSVAFPHNLIDIDPKYIPYKPMNRYNGAMD 636
DB 361 KSVYRALLLNQTLVVDHIERQEDSPINSVAFPHNLIDIDPKYIPYKPMNRYNGAMD 420
QY 637 VMDAHYKMFDFDHGNSPMASIOEAQAEFKRWTFQVNVTFQMESTITRIAYVFGPY 696
DB 421 VMDAHYKMFDFDHGNSPMASIOEAQAEFKRWTFQVNVTFQMESTITRIAYVFGPY 480
QY 697 INVSSCRFIDSSNPGQLQISLVNNTNTEHVVSIVTDYHNLKTRFNLYLGGFGFASVADQGIT 756
DB 481 INVSSCRFIDSSNPGQLQISLVNNTNTEHVVSIVTDYHNLKTRFNLYLGGFGFASVADQGIT 540
QY 757 RFLGLGTAIVKPVHRDRIIPFGKFNIAVGLILCISLVLTQWRFYLSFRKLMRWILI 816
DB 541 RFLGLGTAIVKPVHRDRIIPFGKFNIAVGLILCISLVLTQWRFYLSFRKLMRWILI 600
QY 817 LVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVVITSLPGSGAE 876
DB 601 LVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVVITSLPGSGAE 660
QY 877 ILKQPFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSL 936
DB 661 ILKQPFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSL 720
QY 937 VQDTKLHLQNLHLEPNRGKLAQYFAMNKDKKK 970
DB 721 VQDTKLHLQNLHLEPNRGKLAQYFAMNKDKKK 754

RESULT 8

AA72641
ID AAY72641 standard; protein; 596 AA.
XX
AC AAY72641;
XX
XT 02-MAY-2001 (first entry)
XX Human glycosyl sulfotransferase-6 (GST-6) fragment.
XX Human; glycosyl sulfotransferase-6; GST-6; immunosuppressive; therapy;

KW selectin binding inhibitor; gene therapy; inflammation;
KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW Polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW Glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW asthma; hypersensitivity; rheumatic fever; tissue rejection.

XX Homo sapiens.
OS WO200106015-A1.

XX 25-JAN-2001.

XX 19-JUL-2000; 2000WO-US019741.

XX 20-JUL-1999; 99US-0144694P.

XX 13-JUN-2000; 2000US-00593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI; 2001-138471/14.

XX N-PSDB; AAD02702, AAD02703.

XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for

XX diagnostic and therapeutic agent screening applications.

XX Claim 3; Fig 3; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-6 (GST-6)
CC fragment. GST is a type 2 membrane protein useful for inhibiting a
CC binding event between a selectin and a selectin ligand, which comprises
CC contacting the selectin with a non-sulphated selectin ligand, GST and a
CC small molecular agent that inhibits the sulphation activity of GST. GST
CC is also useful in inhibiting a selectin mediated binding event. GST is
CC useful in gene therapy to treat disorders such as acute or chronic
CC inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis,
CC polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis,
CC diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome,
CC Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism,
CC pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative
CC colitis, dermatitis, myocarditis, regional enteritis, adult respiratory
CC distress syndrome, infantile eczema, psoriasis lichen planus, allergic
CC rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue
CC rejection during transplantation

XX Sequence 596 AA;

Query Match 48.2%; Score 3147; DB 4; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.3e-278;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 MRRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAQAEFKRWTFQVNVTFQMESTIT 686
DB 1 MRRYNGAMMDVMDAHYKMFDFDHGNSPMASIOEAQAEFKRWTFQVNVTFQMESTIT 60

QY 687 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTNTEHVVSIVTDYHNLKTRFNLYLGGFGF 746

DB 61 RIAYVFGPYINVSRCRFDSSNPGQLQISLVNNTNTEHVVSIVTDYHNLKTRFNLYLGGFGF 120

QY 747 ASVADQGITRFLGTAIVKPVHRDRIIPFGKFNIAVGLILCISLVLTQWRFYLS 806

DB 121 ASVADQGITRFLGTAIVKPVHRDRIIPFGKFNIAVGLILCISLVLTQWRFYLS 180

QY 807 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 866

DB 181 FRKLMRWILILVIALWFIELLDVWSTCSQPICAKWTREAGSKSLSEGGHMDLPDVV 240

QY 867 ITSLPGSGAEILKQPFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 926

Db 241 ITSLLPGSABILKQLFNSDFLYIRVPTAYIDIPETELEIDSFDVACEKWKSDIRSGHF 300
QY 927 RLRLGWTQSLVQDTKLHLQNLHLHPNRGKLAQYFANNKDKKFKRRESLPRQSRQMG 986
Db 301 RLRLGWTQSLVQDTKLHLQNLHLHPNRGKLAQYFANNKDKKFKRRESLPRQSRQMG 360
QY 987 AFORDAEYIRALRRHLVYPSARPVLSLSSGSGWTLLKHLFFQEVLGASMRALYIVRDPRAW 1046
Db 361 AFORDAEYIRALRRHLVYPSARPVLSLSSGSGWTLLKHLFFQEVLGASMRALYIVRDPRAW 420
QY 1047 IYMLYNSKPSLYSLKQVPEHLAKLFKEGGKGCNLSNGYAFEBYELRKLSSKSNV 1106
Db 421 IYMLYNSKPSLYSLKQVPEHLAKLFKEGGKGCNLSNGYAFEBYELRKLSSKSNV 480
QY 1107 SLLSHLWLANTAALRINTDLLPTSQYLVKFEDIVHPPOKTERIFAPLGIPLSPASLNQ 1166
Db 481 SLLSHLWLANTAALRINTDLLPTSQYLVKFEDIVHPPOKTERIFAPLGIPLSPASLNQ 540
QY 1167 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 1222
Db 541 ILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKFMD 596
RESULT 9
AAB94410
ID AAB94410 standard; protein; 480 AA.
XX AAB94410;
XX 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:14997.
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EPI074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX Claim 8; SEQ ID NO 14997; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dr primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAB13629 to AAB13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 480 AA;
Query Match 39.8%; Score 2597; DB 4; Length 480;
Best Local Similarity 99.8%; Pred. No. 2.4e-228;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 320 MHFMYATLLPGFQRTVGIADSNYNWFGPESQVFLDKFKILKNGAENWLAQOIRKHP 379
Db 1 MHFMYATLLPGFQRTVGIADSNYNWFGPESQVFLDKFKILKNGAENWLAQOIRKHP 60
QY 380 KQGPMPVSTAQRWSTLHTEYIWDYDQLTPOPPADYGTAKIHTFPNMGVVTYAGLPTQT 439
Db 61 KQGPMPVSTAQRWSTLHTEYIWDYDQLTPOPPADYGTAKIHTFPNMGVVTYAGLPTQT 120
QY 440 NTFVSKGKLGGRVYDIVHPOYSWIDGWRSPNPGHEHPDQNSFTFAPNGOVFVSEAL 499
Db 121 NTFVSKGKLGGRVYDIVHPOYSWIDGWRSPNPGHEHPDQNSFTFAPNGOVFVSEAL 180
QY 500 YGPKLSHLNNLVFAPSPSQCNKPWEGQLGECQAQMLKWTGEEVGDAAAGIITASQHGEM 559
Db 181 YGPKLSHLNNLVFAPSPSQCNKPWEGQLGECQAQMLKWTGEEVGDAAAGIITASQHGEM 240
QY 560 VFVSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVAFPHNLDIF 619
Db 241 VFVSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVAFPHNLDIF 300
QY 620 KYTPYKPMRYNGAMDMVMDAHYKMFWDHGHNSPMASIOEASQAAEFKKRWTFQVNVTF 679
Db 301 KYTPYKPMRYNGAMDMVMDAHYKMFWDHGHNSPMASIOEASQAAEFKKRWTFQVNVTF 360
QY 680 QMESTITRIAYFYGPYINVSRCRFDSSNPGQLISLVNNTTHVSVIVTDYHNLKTRFN 739
Db 361 QMESTITRIAYFYGPYINVSRCRFDSSNPGQLISLVNNTTHVSVIVTDYHNLKTRFN 420
QY 740 YLGGFGFASVADQGITRFGLTQAIKVPVRHRIIPPGFKFNIAVGLILCISLVILTF 799
Db 421 YLGGFGFASVADQGITRFGLTQAIKVPVRHRIIPPGFKFNIAVGLILCISLVILTF 480
RESULT 10
AAY51120
ID AAY51120 standard; protein; 958 AA.
XX AAY51120;
XX AC AAY51120;
XX 24-MAR-2000 (first entry)
XX Human SART-2 protein.
XX SART-2; human; tumor antigen protein; T cell; MHC class I antigen;
XX major histocompatibility complex.
XX Homo sapiens.
XX JPI1318455-A.
XX 24-NOV-1999.
XX 08-MAY-1998; 98JP-00126398.
XX

```

PR 08-MAY-1998; 98JP-00126398.
XX (ITOY/) ITO Y.
XX WPI; 2000-090523/08.
XX DR N-ESDB; AAZ44181.
XX Human cancer involution antigen protein.
XX FT
XX PS Claim 1; Page 11-15; 19pp; Japanese.
XX CC This invention describes a novel human tumor antigen protein (SART-2)
XX CC recognized by T cells by combining with major histocompatibility complex
XX CC (MHC) class I antigens. The tumor antigen protein can be used for the
XX CC treatment and the diagnosis of tumor cells other than melanoma cell,
XX CC particularly canceroid. This sequence represents the human SART-2 protein
XX CC described in the method of the invention,
XX SQ Sequence 958 AA;

Query Match 26.8%; Score 1753.5; DB 3; Length 958;
Best Local Similarity 49.0%; Pred. No. 1.5e-150;
Matches 341; Conservative 114; Mismatches 210; Indels 31; Gaps 9;

QY 72 HPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKHADFAAKWNEYI 131
DB 43 HPMYFSAEVAELQLRASASHEHIAARLTAHVHTMLSSPLEYLPWDPKDYSAKWEIF 102
QY 132 GNNLPPLALYCLCPEDKVAFEFVLEVMRVMGVKDWLVENAFGDEVPIGHSITGFATF 191
DB 103 GNNLGALAMFCVLYPENIEARDMAKDYMERMAAQPSMLVKDAPWDEVPLAHSVLGFATY 162
QY 192 DFLYNLLDNHRRQKYLEKIWIITEEMVEYSKVRSGKQLLNHQAATNMIALTGALVTGV 251
DB 163 DFLYNLSKTQOEKFELEIVANASGYMYETSYRRGWFQYLLHHQPTNCWALLTGSVL-M 221
QY 252 DKG--SKANITWQAVDVMKTFPLNHIVDGSLDEGVAYGVTAKSVTOYVFLAQRFN 309
DB 222 NQGVLOEAYLWTKQVLTIMESLVLLREVTDGSLYEGVAGSVTTSLSLFQYMLVQRHN 281
QY 310 INNDNNLWKHFYFATLLPGFQRTVGADSNYNWFYSPESQVLPDLKFLKNGAGNW 369
DB 282 INHGFHPLKQHFAMFYRTLLPGFQRTVAIDSNYNWFYSPESQVLPDLKFLKNGAGNW 341
QY 370 LAQQRKRPKDGMPVSTQARWSTLHTEYVWDPLTPPPADYGTAKHTHPNNGVVT 429
DB 342 LADQIRNRVVEGPTSGKGRWCTLHTEFLWYDGLSKSVPPDPFGPTLHYFEDWGVVT 401
QY 430 YGAGLPNTQNTFVSFKSGKLGRAVDIVHFQPY-SWIDGWSFNPGEHPDQNSFTFA 488
DB 402 YGSALPAEINRSFLFSKGLGGRALVDIVRNKYKDWIKGNWRFNAGHEHPDQNSFTFA 461
QY 489 PNGQVFSVEALYGPKLHNLVLPSPSSQCNKPWEGQLGE--CAQMLKWTGVEVGDA 546
DB 462 PNGVPFTEALYGPKYTFNNVLMFSPAVSKSCFSPWVGQVTDCCSKSKYKHDLAASC 521
QY 547 AGEIITASQHEGMVFSVGEASVAYSSAMRLKSVYRALLILNSQTLLVWDHIERQEDSPIN 606
DB 522 QGRVVAEEKNGVVFIIRGEGVAGYNPOLNKLQVNRNLIHLHPQLLLVQIHLGEESPLE 581
QY 607 SVSAFFHNLIDFKYIPKFMNRYNGAMDMVDVAHYKMFWDHSGNSPMASIQEAQAAE 666
DB 582 TAASFFHNVDPPE--ETVVDGVGHAFIRQRGLYKMYWMDTGTSEKATFASVYTPRG 638
QY 667 FKRWTOFVNVTQEMESTITRIAYFVGPYINVSRCRFDSSNPGIQLISLNVNTHVVS 726
DB 639 YPVGNTNVNVTMHLRSPITRAAYLFIGPSIDYQS-----FTVHGDSQQLDVF 686
QY 727 IVDYHNLKTRFNVL-----GRGGFASV-ADQGOI 755
DB 687 IATSKHAYAT---YLTWTEATGQSFAAQVADRHKI 719

```

```

RESULT 11
ADL83102
ID ADL83102 standard; protein; 958 AA.
XX AC ADL83102;
XX DT 17-JUN-2004 (first entry)
XX DE Human PRO49564, SEQ ID 304.
XX KW Immunosuppressive; Cytostatic; Antiarthritic; Antirheumatic; Antianemic;
XX KW Antiallergic; Muscular; Neuroprotective; Nephrotropic; Antiinflammatory;
XX KW Gene Therapy; PRO; B cell related disorder; cancer;
XX KW immune-mediated inflammatory disease; human.
XX OS Homo sapiens.
XX PN WO2004024097-A2.
XX PD 25-MAR-2004.
XX PF 15-SEP-2003; 2003WO-US029097.
XX PR 16-SEP-2002; 2002US-0411392P.
XX (GETH ) GENENTECH INC.
XX PI Chiu H, Clark H, Dennis K, Pong S, Schoenfeld JR, Wood WI;
XX PU Wu TD;
XX WPI; 2004-329389/30.
XX DR N-PSDB; ADL83101.
XX New PRO polypeptide, useful for diagnosing and treating a B cell related
PT disorder, e.g. Burkitt's lymphoma, rheumatoid arthritis, autoimmune
PT mediated hemolytic anemia, myasthenia gravis or ankylosing spondylitis.
XX Claim 10; Fig 304; 695pp; English.
XX The present invention relates to PRO proteins and their coding sequences.
XX The PRO proteins are useful for diagnosing and treating a B cell related
CC disorder, e.g. X-linked infantile hypogammaglobulinemia, polysaccharide
CC antigen unresponsiveness, selective IgA deficiency, selective IgM
CC deficiency, selective deficiency of IgG subclasses, immunodeficiency with
CC hyper IgM, transient hypogammaglobulinemia of infancy, Burkitt's
CC lymphoma, intermediate lymphoma, follicular lymphoma, type II
CC hypersensitivity, rheumatoid arthritis, autoimmune mediated haemolytic
CC anaemia, myasthenia gravis, hypoadrenocorticism, glomerulonephritis, or
CC ankylosing spondylitis. The PRO proteins are also useful for preparing a
CC medicament for treating a condition that is responsive to the PRO
CC protein, e.g. cancer or immune-mediated inflammatory diseases. The PRO
CC coding sequences are useful as hybridization probes in chromosome and
CC gene mapping, in preparing PRO proteins, or in generating transgenic
CC animals or knockout animals, which in turn are useful in the development
CC and screening of therapeutically useful reagents.
XX SQ Sequence 958 AA;

Query Match 26.8%; Score 1753.5; DB 8; Length 958;
Best Local Similarity 49.0%; Pred. No. 1.5e-150;
Matches 341; Conservative 114; Mismatches 210; Indels 31; Gaps 9;

QY 72 HPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKHADFAAKWNEYI 131
DB 43 HPMYFSAEVAELQLRASASHEHIAARLTAHVHTMLSSPLEYLPWDPKDYSAKWEIF 102
QY 132 GNNLPPLALYCLCPEDKVAFEFVLEVMRVMGVKDWLVENAFGDEVPIGHSITGFATF 191
DB 103 GNNLGALAMFCVLYPENIEARDMAKDYMERMAAQPSMLVKDAPWDEVPLAHSVLGFATY 162
QY 192 DFLYNLLDNHRRQKYLEKIWIITEEMVEYSKVRSGKQLLNHQAATNMIALTGALVTGV 251
DB 163 DFLYNLSKTQOEKFELEIVANASGYMYETSYRRGWFQYLLHHQPTNCWALLTGSVL-M 221

```


QY 601 EDSPINSVSAFFHNLIDIDFKYIPYKFMRYNGAMMDVDAHYKMFWDHGHGNSPMASIOE 660
Db 197 -----196
QY 661 AEQAAEFKRWTQFVNVTFQWSTITRIAYVFGPYINVSCEFIDSSNPGQLQISLNVN 720
Db 197 -----196
QY 721 TEHVVSIVTDYHNLKTRFNVLGRGFASVADQQTFRFGLGTQAIKVPKVRHDIRIIFPGF 780
Db 197 -----196
QY 781 KFNIAVGLILCISLVILTFQWRFLSPKRLMRWILIIIVIALWFIELDVWSTCSQPICAK 840
Db 197 -----196
QY 841 WTRTEAGSKSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDI 900
Db 197 -----CKKSLSEGHMDLPDVVITSLPGSGAEILKOLFNFSSDFLYIRVPTAYIDI 248
QY 901 PETELEIDSVDACEWKVSDIRSGHFRLLRCWLQSLVODTKLHLQNIHLHEPNRGKLAQY 960
Db 249 PETELEIDSVDACEWKVSDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHEPNRGKLAQY 308
QY 961 FANWKD 966
Db 309 FANWKD 314
RESULT 13
ADRI0287
ID ADR10287 standard; protein; 625 AA.
XX AC ADR10287;
XX DT 04-NOV-2004 (first entry)
XX DE Human protein useful for treating neurological disease Seq 3793.
XX KW human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; neurotropic; antiparkinsonian; cytosstatic;
KW tranquiliser.
XX OS Homo sapiens.
XX PN EPI447413-A2.
XX PD 18-AUG-2004.
XX PF 12-FEB-2004; 2004EP-00003145.
XX PR 14-FEB-2003; 2003JP-00102207.
XX PR 09-MAY-2003; 2003JP-00131452.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX PI Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX WPI; 2004-583265/57.
DR N-PSDB; ADR08331.
XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's disease, dementia and various cancers.
XX Claim 1; SEQ ID NO 3793; 2686pp; English.
XX This invention relates to novel, isolated full length human cDNA
CC molecules and the encoded proteins thereof. Specifically, it refers to

CC cDNA clones obtained by an oligo-capping method, where none of these
CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, neurotropic, antiparkinsonian,
CC cytosstatic and tranquiliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.
SQ Sequence 625 AA;
Query Match 13.7%; Score 898; DB 8; Length 625;
Best Local Similarity 44.6%; Pred. No. 2e-72; Mismatches 131; Indels 28; Gaps 7;
Matches 180; Conservative 65;
QY 362 LKNGAGNMLAQIIRKRPKDGMPVPSTAQRMSTLHTEIYWDPLQTPQADYGTAKIHT 421
Db 1 MRNGSGNWLADQIRNRVVEGPGTSGKQRMCTLHTEFLWYDGLSKSVPPDPFGPTLHY 60
QY 422 FPNWGVVYTGAGLPNTQNTNTFVSFKSGKLGRAVVDIVHFQPY-SWIDGWSRPNRGHEHP 480
Db 61 FEDWGVVYTGSAIPAEINRSFLSKGKLGRAIYDIVHRNKYKDWIRKGRNFAHGEHP 120
QY 481 DONSFTFAPNGQVFSVSEALYGPKLHLNNVLVAFSPSSQCNKPWEGOLGE--CAQWLKW 538
Db 121 DONSFTFAPNGVFPFTEALYGPYTFNNVLMFSPAVSKSCFSPVWGVQVTEDCSKWSKY 180
QY 539 TGEVGDAAAGEIITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLNLSQTLVVVDHIE 598
Db 181 KHDLAASCQGRVAAEKEKNGVVFIRGEGVAYNPOLNKNVQRLNLLHPOLLVLDQIH 240
QY 599 RQEDSPINSVSAFFHNLIDIDFKYIPYKFMRYNGAMMDVDAHYKMFWDHGHGNSPMASI 658
Db 241 LGEESPLETAASFFHNVDVPFE--ETVVDGVHGAFIRQRDLGYKMYWMDDTGYSEKATF 297
QY 659 QEAQAAEFKRWTFQVNVTFQWSTITRIAYVFGPYINVSCEFIDSSNPGQLQISLNV 718
Db 298 ASVTYPRGYPYNGTNYVNVMTMHLRSPITRAAYLFTGPSIDVQS-----FTVHG 345
QY 719 NNTHEWVSIVTDYHNLKTRFNYL-----GFGGFASV-ADQOI 755
Db 346 DSQQLDVFIAATSKHAYT---YLTWGEATGQSFAQVIADRHKI 386
RESULT 14
ADG27796
ID ADG27796 standard; protein; 128 AA.
XX AC ADG27796;
XX DT 26-FEB-2004 (first entry)
XX DE Human novel protein amino acid sequence SeqID563.
XX KW antiparkinsonian; haemostatic; neurotropic; neuroprotective; osteopathic;
KW anti-HIV; protozoacide; antifungal; immunosuppressive; antineumatic;
KW antiarthritic; antidiabetic; anti allergic; antiinflammatory;
KW anticoagulant; cytosstatic; gene therapy; Parkinson's disease;
KW Alzheimer's disease; thrombocytopaenia; osteoporosis; osteoarthritis;
KW infection; HIV; Leishmania; malaria; fungal infection;
KW multiple sclerosis; rheumatoid arthritis;
KW insulin dependent diabetes mellitus; allergic reaction; food allergy;
KW insect allergy; allergic rhinitis; haemophilia; cancer; human.
XX

OS Homo sapiens.
 XX WO200179254-A1.
 XX
 XX
 XX
 XX 25-OCT-2001.
 XX
 XX 16-APR-2001; 2001WO-US008655.
 XX
 XX 18-APR-2000; 2000US-00552929.
 XX 22-SEP-2000; 2000US-00668317.
 XX 24-OCT-2000; 2000US-00695783.
 XX 01-DEC-2000; 2000US-00728628.
 XX 26-JAN-2001; 2001US-00770160.
 XX 13-FEB-2001; 2001US-00783066.
 XX 22-MAR-2001; 2001US-00816928.
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Asundi V, Zhou P, Xue AJ, Ren F, Zhang J, Wang J;
 PI Yang Y, Zhao QA, Goodrich RW, Liu C, Drmanac RT, Ma Y, Wang Z;
 PI Wehrman T;
 XX
 XX WPI; 2001-607699/69.
 XX
 XX Novel polynucleotides and encoded polypeptides (protein factors,
 PT including, e.g., cytokines, such as lymphokines, interferons, and
 PT circulating soluble factors) useful for treating, e.g., Parkinson's,
 PT Alzheimer's, HIV and cancer.
 XX
 XX Claim 20; SEQ ID NO 563; 153pp; English.
 XX
 XX This invention relates to a novel isolated DNA sequence and the mature
 CC proteins encoded by them. The invention may be useful in the development
 CC of compositions with antiparkinsonian, haemostatic, nootropic,
 CC neuroprotective, osteopathic, anti-HIV, protozoacide, antifungal,
 CC immunosuppressive, antiarthritis, antiarthritic, antidiabetic,
 CC antiallergic, antiinflammatory, anticoagulant or cytostatic activities.
 CC In addition, the sequences of the invention may be useful for gene
 CC therapy. The invention may be useful for the development of treatments
 CC for Parkinson's, Alzheimer's, thrombocytopaenia, osteoporosis, and
 CC osteoarthritis, infections (including HIV, Leishmania, malaria, and
 CC various fungal infection), autoimmune disorders such as multiple
 CC sclerosis, rheumatoid arthritis, and insulin dependent diabetes mellitus,
 CC allergic reactions and conditions (for example food allergies, insect
 CC allergies and allergic rhinitis), coagulation disorders including
 CC haemophilia, and cancer. Note: The amino acid sequences given in table 6
 CC (SeqID 439-584) may have in frame stop codons or possible
 CC insertions/deletions as shown in the table. The sequences allocated Seq
 CC IDs 1-438 are not provided (even by reference) in the specification.
 XX
 XX Sequence 128 AA;
 XX
 XX Query Match 9.8%; Score 637; DB 4; Length 128;
 XX Best Local Similarity 96.1%; Pred. No. 1.2e-49;
 XX Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Qy 1096 KELSXSNAVSLASHLWLANTRAALRINTDLLPTSQVLKFEIVHPQKTERIAFL 1155
 Db 2 KNYQPKSNVSLASHLWLANTRAALRINTDLLPTSQVLKFEIVHPQKTERIAFL 61
 Qy 1156 GIPLSPASNLQILFATSNLFLPYVEGEISPTNTNVWQKQNLPROEIKLIENICWTLMDRL 1215
 Db 62 GIPLSPASNLQILFATSNLFLPYVEGEISPTNTNVWQKQNLPROEIKLIENICWTLMDRL 121
 Qy 1216 GYPKFMD 1222
 Db 122 GYPKFMD 128
 RESULT 15
 ADJ12217
 ID ADJ12217 standard; protein; 125 AA.
 XX

AC ADJ12217;
 XX
 XX 20-MAY-2004 (first entry)
 XX
 XX Human secreted protein SeqID 71.
 XX
 XX human; secreted; cancer; haematopoietic disease; anaemia;
 KW multiple myeloma; reproductive system disorder; prostatitis;
 KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;
 KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;
 KW fetal alcohol syndrome; Down's syndrome; excretory disease;
 KW urinary incontinence; renal disorder; neural; sensory disease;
 KW Alzheimer's disease; meningitis; respiratory disease; emphysema;
 KW occupational lung disease; endocrine disease; portal hypertension;
 KW glomerulonephritis; digestive disease; scleroderma;
 KW irritable bowel syndrome; epithelial disease; antiarthritic;
 KW epidermolysis bullosa; cytostatic; antianemic; antiinflammatory;
 KW antiasthmatic; anti-HIV; immunosuppressive; antiarthritic;
 KW antipsoriatic; antibacterial; osteopathic; dermatological; antigout;
 KW immunomodulator; antiarrhythmic; cardiant; nootropic; antilipemic;
 KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
 KW antidiabetic; anabolic; hypertensive; vulnery.
 XX
 XX Homo sapiens.
 OS
 XX US2004010132-A1.
 XX
 XX 15-JAN-2004.
 PD
 XX
 XX 30-OCT-2001; 2001US-00984429.
 PF
 XX
 XX 09-OCT-1997; 97US-0061463P.
 PR
 XX 09-OCT-1997; 97US-0061527P.
 PR
 XX 09-OCT-1997; 97US-0061529P.
 PR
 XX 09-OCT-1997; 97US-0061532P.
 PR
 XX 09-OCT-1997; 97US-0061536P.
 PR
 XX 09-OCT-1997; 97US-0071498P.
 PR
 XX 08-OCT-1998; 98MO-US021142.
 PR
 XX 08-APR-1999; 99US-00288143.
 PR
 XX 01-NOV-2000; 2000US-0244591P.
 PR
 XX (ROSE/) ROSEN C A.
 PA (BREW/) BREWER L A.
 PA (DUAN/) DUAN R D.
 PA (RUBE/) RUBEN S M.
 PA (FLOR/) FLORENCE K A.
 PA (GREE/) GREENE J M.
 PA (YOUN/) YOUNG P E.
 PA (FERR/) FERRIE A M.
 PA (YUGG/) YU G.
 PA (FLOR/) FLORENCE C.
 PA (EBNE/) EBNER R.
 PA (OLSE/) OLSEN H.
 XX
 XX Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM;
 PI Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;
 XX
 XX WPI; 2004-090518/09.
 DR N-PSDB; ADJ12161.
 DR
 XX New isolated nucleic acids and polypeptides, useful for diagnosing,
 PT treating, preventing or ameliorating diseases or disorders e.g. cancer,
 PT anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's
 PT disease.
 XX
 XX Claim 11; SEQ ID NO 71; 286pp; English.
 XX
 XX This invention relates to novel polynucleotides encoding human secreted
 CC proteins. Specifically, it refers to the vectors, host cells, recombinant
 CC and synthetic methods for producing human polynucleotides, polypeptides
 CC and antibodies. Furthermore, it relates to screening methods to identify
 CC agonists and antagonists that can be used to inhibit or enhance the
 CC production and function of the secreted proteins. The present invention

describes these compositions as useful for diagnosing, treating or preventing disorders such as cancer, haematopoietic diseases including anaemia and multiple myeloma, reproductive system disorders including prostaticitis and inguinal hernia, musculoskeletal diseases including systemic lupus erythematosus and gout, cardiovascular disease including arrhythmia and hypernatraemia, mixed fetal diseases including fetal alcohol syndrome and Down's syndrome, excretory diseases including urinary incontinence and renal disorders, neural or sensory disease including Alzheimer's disease and meningitis, respiratory disease including emphysema and occupational lung disease, endocrine diseases including diabetes and glomerulonephritis, digestive diseases including portal hypertension and irritable bowel syndrome and connective tissue or epithelial diseases including scleroderma and epidermolysis bullosa. As such, there are various activities such as cytostatic, antianemic, antiarthritic, antiasthmatic, anti-HIV immunosuppressive, antiinflammatory, antipsoriatic, antibacterial, osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic, cardiant, neurotropic, antilipemic, nephrotropic, uropathic, neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and vulnervary. This polypeptide is a human secreted protein of the invention. NOTE: This sequence does not appear in the printed specification but has been obtained in electronic format from the US patent office at the following web site www.seqdata.uspto.gov/sequence.html; Document ID: 20040010132.

SQ Sequence 125 AA;

Query Match 8.7%; Score 568; DB 8; Length 125;
Best Local Similarity 100.0%; Pred. No. 2.5e-43;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	811	MRWILLVIALWFIELDDVWSTCSQPTCAKWTTEAGSKLSLSGGHMDLPDVVITSL	870
Db	1	MRWILLVIALWFIELDDVWSTCSQPTCAKWTTEAGSKLSLSGGHMDLPDVVITSL	60
Qy	871	PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELIDSFVDACEWK	917
Db	61	PGSGAEILKQLFFNSSDFLYIRVPTAYIDIPETELIDSFVDACEWK	107

Search completed: June 23, 2005, 08:43:30
Job time : 135.064 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:34:49 ; Search time 46.0594 Seconds
(without alignments)

1980.512 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWTIMALMFTGHL.....LIENICWTLMDRLGYPKFM 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pbp:*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pbp:*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pbp:*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pbp:*
5: /cgn2_6/ptodata/1/iaa/PCUS COMB.pbp:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pbp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568	8.7	126	4	US-09-288-143-71
2	147	2.3	411	3	US-09-015-188-2
3	145	2.2	2504	4	US-08-851-567B-12
4	145	2.2	2504	4	US-09-817-514A-8
5	135	2.1	1844	4	US-08-851-567B-53
6	134	2.1	483	3	US-09-263-023-2
7	134	2.1	483	4	US-09-471-867-2
8	133	2.0	484	3	US-09-263-023-4
9	133	2.0	484	4	US-09-471-867-4
10	133	2.0	531	4	US-09-949-016-6471
11	133	2.0	608	4	US-09-949-016-9449
12	127	1.9	1456	4	US-09-976-594-168
13	123.5	1.9	395	4	US-09-949-016-7011
14	120.5	1.8	479	2	US-08-899-514-2
15	119.5	1.8	390	4	US-09-949-016-6813
16	119.5	1.8	431	4	US-09-949-016-8893
17	115.5	1.8	386	4	US-09-786-240-11
18	115	1.8	893	3	US-09-514-302-4
19	115	1.8	893	4	US-10-014-436-4
20	115	1.8	1938	3	US-09-514-302-2
21	115	1.8	1938	4	US-10-014-436-2
22	112	1.7	772	1	US-08-258-639A-2
23	112	1.7	772	2	US-08-900-951-2
24	112	1.7	772	5	PCT-US95-07391A-2
25	111.5	1.7	462	2	US-08-870-180B-13
26	111.5	1.7	462	3	US-09-226-529-13
27	110.5	1.7	1455	3	US-08-840-062-5

28	108.5	1.7	386	3	US-09-045-284A-2	Sequence 2, Appli
29	108.5	1.7	386	3	US-09-190-911-1	Sequence 1, Appli
30	108	1.7	826	4	US-09-252-991A-22143	Sequence 22143, A
31	105	1.6	391	4	US-09-248-796A-15903	Sequence 15903, A
32	105	1.6	1180	3	US-09-224-024-28	Sequence 28, Appl
33	105	1.6	1180	5	PCT-US94-07902-28	Sequence 28, Appl
34	104.5	1.6	843	4	US-09-248-796A-17210	Sequence 17210, A
35	104	1.6	1259	3	US-09-134-001C-3757	Sequence 3757, Ap
36	104	1.6	1504	4	US-09-676-519-27	Sequence 27, Appl
37	103.5	1.6	324	4	US-09-270-767-43224	Sequence 43224, A
38	103	1.6	582	3	US-09-091-725-17	Sequence 17, Appl
39	102	1.6	1227	4	US-09-710-279-96	Sequence 96, Appl
40	101.5	1.6	594	3	US-08-987-151-2	Sequence 2, Appli
41	101	1.5	574	4	US-09-248-796A-18722	Sequence 18722, A
42	101	1.5	650	4	US-09-248-796A-16387	Sequence 16387, A
43	100.5	1.5	1172	4	US-09-328-352-6071	Sequence 6071, Ap
44	100	1.5	630	4	US-09-538-092-384	Sequence 384, App
45	100	1.5	906	4	US-09-417-039-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1

US-09-288-143-71
; Sequence 71, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,463
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,527
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,536
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,532
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
; US-09-288-143-71

Query Match 8.7%; Score 568; DB 4; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.8e-49;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	811	MRVILLIVIALWTFELDVWSTCSQPCAKWTEARSGKSLSEGHMDLPDVTLSL	870
Db	1	MRVILLIVIALWTFELDVWSTCSQPCAKWTEARSGKSLSEGHMDLPDVTLSL	60
Qy	871	PGSGAELTKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK	917
Db	61	PGSGAELTKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWK	107

RESULT 2

```
US-09-015-188-2
; Sequence 2, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 411
; TYPE: PRN
; ORGANISM: Homo sapiens
; US-09-015-188-2

Query Match      2.3%; Score 147; DB 3; Length 411;
Best Local Similarity 18.2%; Pred. No. 1.9e-05;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY      813 WILVIALWFIEL-----LDVMSTC-----SQPICAKWTRTEAGSKSLSS 855
DB      5 WKAVLLALASIAIQYTAIRTFKSFHTCTGLAEAGLAERLC-----ESPTFAYNLSR 59

QY      856 EGHMDLPDVVITSLPGSGAIEIKQLFFNSSDFLYTRVPTAYIDIPETELEIDSFVDACE 915
DB      60 KTH-----ILILATRRSGSFGQLFNQHLDFVLFEP-----93

QY      916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHBPNGKLAQYFAMNKDKKPKFRE 975
DB      94 -----YHVNTLIPRTQG-----KSPADRV 115

QY      976 SLPEQRSMKGAFDRDAEYI-----RALRRHLVYYPGARPV-----LS 1013
DB      116 MCGASRDLLSLYDCDLYFLENYKPPVNHVTDTRFRGASRVLCSPVCDPPGADLV 175

QY      1014 LSSGSWTLKHFFQEVUGA-----SMRALYIVRD 1042
DB      176 LEEGDCVRKCGLLNLTVAAEACRSHVAIKTVRVPVNDLRALVEDPRLNLKVIQLVRD 235

QY      1043 PRAWISMLYNSKPSLYSLKNVPEHLAKLPKIEGGKGNLNSGYAFEYELRKLKSK 1102
DB      236 PRGILASRSTFRDY-----RLRLWYGTGR-----KPYNLDTQL-TTVCEDF 279

QY      1103 SNAVS--LLSHLWLANTAALRINTDLLPTSQYLVKPEDIVHPPOKTTTERRIFAFGLIPL- 1159
DB      280 SNSVSTGLMRPPLKG-----KYLVRVEDLARPNMKKTEIYGFGLIPLD 325

QY      1160 -----SPASLNQILFATSNLFLVFLPYEGEISPTNTNVWKNLPDRBEIKLIENI 1207
DB      326 SHVARMTQNTNRGDTPLGKHGYTVRN-----SAATAEKWRFLSYDIVAPAQNA 375

QY      1208 CWTLMDRLG 1217
DB      376 CQVLAQLGY 385

RESULT 3
US-08-851-567B-12
; Sequence 12, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: Ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
```

```
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-851-567B-12

Query Match      2.2%; Score 145; DB 4; Length 2504;
Best Local Similarity 18.1%; Pred. No. 0.00091;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

QY      29 FSTFEESVSNYSEWA-----VFTDDIDQFKTKQVDFRPNQKLKSLMHPSLYFD 78
DB      1003 FTDWERYNKRYSTWAGVSELVYYPENYVDPTQRIGQTKQMD-ALLQSIQSLQADNTYED 1061

QY      79 AGEIOAMQKSRASHLHLFRAIRSAVTV-----MLSNPTYLPPPKHA-----D 122
DB      1062 AFKTY-LTSFEQVANLKVISAYHDNVNVDQGLTYFIDQAAPGTYYWRSVDHSCENGK 1120

QY      123 FAA-----KQNEIYGNLPLALYCLLCPEDKVAPEFVLEVMYKDMVGVKDLVENAPGDEV 178
DB      1121 FAANAWGEWNKI-----TCAVNPWKNIIRPVV--YMSRL--YLLMLEQQSKKSD- 1165

QY      179 PIGHSLTGFAFAFDPLYNLNDHRRQKYLEKIW-----VITEEMYEYSKVRSMGKQLLH 232
```

Db 1166 -----DQKTTIYQ--YNLKLHRI--YDGSMTPTFDVTEKVNYSSTDAESL-- 1212
Qy 233 NHQATNMIALTGALVTGVDKGSKANIWQAVDVMEKTMFLNHHIVDGSLEGVAGSY 292
Db 1213 -----GLYCTGYQ-----EDTLVMFYMQSS-----YSS 1239
Qy 293 T---AKSVTQVFLAQRHFNINL--NNWLMKHFYATLLPGFQRTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSDDMTNAQATYNNYS-----PQFDTVMADPDS-- 1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQIRKHPKDGMPVPSTAQRWSTLHTIYIYDPLT 407
Db 1287 -----NKKVITRRVNNRYAEDYE-----IPSSV-----TSNSNYSWGDHSLT 1323
Qy 408 POPPADYGTAKIHTFPNMGVVTYAGLPN-----TQNTFVS--FKSGKLGGRV 455
Db 1324 -----MLYGSVPNITFESAEDLRLSTNNALSIHNGVAGTRRI 1363
Qy 456 YDIVHFPYSWIDGRSPNPGHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLPAP 515
Db 1364 QCNLMKQVYASLGDFIYDSSFD--DANRFLNP--LF-----KFGKDESDSICIYNE 1414
Qy 516 SPSSQCNKPWEGQGECAQMLKWTGE-----EVGDAAGEIITASQHGEMVFSVGEAVSAYS 571
Db 1415 NPSSE-DKKWYFSSKDDNKTDYNGTQCIDAGTSNKDFYNNLQIEIEIVISVTGGWSSYK 1473
Qy 572 SAMBLKSVYRALLLNLSQTLVVDHIERQEDSPINSVSAFPHNLDIDPKYIPYKPMRYN 631
Db 1474 IS-----NPKINI-----NTGIDSARKV--KVTVKAG 1496
Qy 632 GAMDMVDHAHYKMFDFHHGNSPMASIQEABQAEFKKRWTFQVNVTFQEST--ITRIA 689
Db 1497 G-----DQIFTDNSTYVQ-----QAPSPFEMIQFNLTIDCKNLNFIHQ 1542
Qy 690 YVFGYPIYINSSCRPIDSSNPGLOISLNNVNTHEVSVITDYHNL-----KTRFNYL 741
Db 1543 HIEIDFTAQDGRFLGAEFTIIPVKVGLGTENVIALYSENNGVQVMOIGAYRTRNL 1602
Qy 742 GFGFASVADOG-----QITRFGLTQAIKVPVRHRIIPFGPKFIAGVLI 789
Db 1603 FAQQLVSRANRGIDAVLSMETQNTQEPQLGAGTVVQLVLDKYDESIGHTNKSFAIEYVDI 1662
Qy 790 LCISAVILTQWRFPYLSPRKLMRWILLVIALWPIELL-----DVMSTCSQPICAKW----- 841
Db 1663 FKENDSFVIYQELSETSQTVK-----VFLSYPIEATGNKNHLM-----VRAKYQKET 1711
Qy 842 -----TRTEAS-----GSKKLSSEGHMDLPDVVITSPLPGSA--- 875
Db 1712 TDKILFDRDEKPHGWFLSDHKTFTSGLSSAQAALKNDSEPMDF-----SGANAL 1761
Qy 876 -----EILKQLFFNSD--FLYIRVPTAYIIDIPELEIDSFDVACEWKV 918
Db 1762 YFWELFYTPMMAHRLLEQNFDAANHFVYMSPSGYI-----VDGKIAIYHNV 1813
Qy 919 SDIRSGHFRLLRGWQLSVODTKLHLNHLHEPN-----RGKLAQYFA-----MN 964
Db 1814 -----RPLEEDTSMNAQQDSDTPDAVAQDDPMHYKVATFMATLDDLMA 1858
Qy 965 KDKKKFKRRRSLPQRSQMKGAFFDRDAEYIRALR-----RHLVYPSARPVLSSSGS 1018
Db 1859 GDAAYRQLERDTLAEAKW-----YTOALNLLGDEPQVMLSTTWANPTLGNAAK 1908
Qy 1019 WTLKHLFTQEVULGASMRALYIVRDPRAWIYMLYNSKPSLSYLVKVPHEHLAKLFIEGK 1078
Db 1909 TQOQVR--QOVL--TQRLNSRVKTP-----LLGTANSLAL--FLPOENSKL-----K 1951
Qy 1079 G-----KCNLSGAFYEPYPRKELSKSNVAVLSLHMLWANTAAALRINTDLLPT 1130
Db 1952 GWRTLAORMENLRNLSIDGQPLSLPYAKPADPKALLSAVSAQSGA----- 2001
Qy 1131 SYQLVKFEDIVH--FPQ 1145

Db 2002 --DLPKAPLTIHRFPQ 2015
RESULT 4
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817.514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-8
Query Match 2.2% Score 145; DB 4; Length 2504;
Best Local Similarity 18.1% Pred. No. 0.00091;
Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;
Qy 29 FSTFEESVSNYSEWA-----VFTDIDQPKTKQVDFRPNQKLKKSMLHPSLYFD 78
Db 1003 FTDWERYNKYSTWAGVSELVYYPENYVDPTRIGQTKMMD-ALLQSIQSLNADTVED 1061
Qy 79 AGEIQAMRQKSRASHLHLFRAIRSAVTV-----MLSNYTYLPPPKHA-----D 122
Db 1062 AFKTY-LTSFEQVANLKVISAYHDNVNVDQGLTVFIGIDQAAPGYVRSVDHSCENGK 1120
Qy 123 FAA-----KMEIYGNLPLALYCLCPEDKVAPEFVLEVMYDMRVGYKDWLVENAPGDEV 178
Db 1121 FAANAWENKNI-----TCAVNPKNIIIPVV--YMSRL--YLLMLEQSKSD- 1165
Qy 179 PIGHSLTGPFATADFLYNLDNHRQKYLEKIW-----VITEWYKSVRSWCKQLLH 232
Db 1166 -----DGKTTIYQ--YNLKLHRI--YDGSMTPTFDVTEKVNYSSTDAESL-- 1212
Qy 233 NHQATNMIALTGALVTGVDKGSKANIWQAVDVMEKTMFLNHHIVDGSLEGVAGSY 292
Db 1213 -----GLYCTGYQ-----EDTLVMFYMQSS-----YSS 1239
Qy 293 T---AKSVTQVFLAQRHFNINL--NNWLMKHFYATLLPGFQRTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSDDMTNAQATYNNYS-----PQFDTVMADPDS-- 1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQIRKHPKDGMPVPSTAQRWSTLHTIYIYDPLT 407
Db 1287 -----NKKVITRRVNNRYAEDYE-----IPSSV-----TSNSNYSWGDHSLT 1323
Qy 408 POPPADYGTAKIHTFPNMGVVTYAGLPN-----TQNTFVS--FKSGKLGGRV 455
Db 1324 -----MLYGSVPNITFESAEDLRLSTNNALSIHNGVAGTRRI 1363
Qy 456 YDIVHFPYSWIDGRSPNPGHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLPAP 515
Db 1364 QCNLMKQVYASLGDFIYDSSFD--DANRFLNP--LF-----KFGKDESDSICIYNE 1414
Qy 516 SPSSQCNKPWEGQGECAQMLKWTGE-----EVGDAAGEIITASQHGEMVFSVGEAVSAYS 571
Db 1415 NPSSE-DKKWYFSSKDDNKTDYNGTQCIDAGTSNKDFYNNLQIEIEIVISVTGGWSSYK 1473
Qy 572 SAMBLKSVYRALLLNLSQTLVVDHIERQEDSPINSVSAFPHNLDIDPKYIPYKPMRYN 631
Db 1474 IS-----NPKINI-----NTGIDSARKV--KVTVKAG 1496

QY 632 GAMDMVMDAHYKMFWDHHCNGSPMAISQEAQAEKFKRWTFQVNVTFQMEST--ITRIA 689
Db 1497 G-----DDQIFDANSYVFO-----QAPSPSEMIYQFNLTIDCKNLNFDNQ 1542
QY 690 YVFGPYINSSCRFIDSSNGPLQISLVNNTTHVSVITDYHNL-----KTRFNYL 741
Db 1543 HIEIDFATAQDGRFLGAETFIIPVTKVLGTENVIALYSENGVQVMOIGAVRTRLNLT 1602
QY 742 GFGFASVADQG-----QITRFLGLTQAIKVPVRHRIIPPGFKFNIAVGLI 789
Db 1603 FAQQLVSRANRGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGNTKNSFAIYVDI 1662
QY 790 LCISLVLTITQWRFLSKMLMILVIALFIELL-----DWSCTSQPICAKW--- 841
Db 1663 FKENDSVIYQGLSETSQTVK-----VFLSFIEATGNKWHM-----VRAKTOKET 1711
QY 842 -----TRTBAE-----GSKSLSEGHMHMDLPDVVITSLPGSGA--- 875
Db 1712 TDKILPRTDEKDPHGWFSLDDHKTFSLSSAQAALKNDSEPMDF-----SCANAL 1761
QY 876 -----ELKQLFNSSD--FLYTRVPTAYIDIPETELEIDSFVDACEWKV 918
Db 1762 YFMELFYTPMMAHRLLOBNQFNDAAHWPYVWSPSGYI-----VDGKIATYHNV 1813
QY 919 SDIRSGHFRLLRGWLQSLVODTKLHLQNIHLHPN-----RGKLAOVFA-----MN 964
Db 1814 -----RLEEDTSNAQQLDSTDPDAVQDDPHYKVATFMATLDLLMAR 1858
QY 965 KDKRKRFRRESLPEQSRQMGKAFDRDAEYIRALR-----RHLVYPSARPVLSSGS 1018
Db 1859 GDAAYRQLERDTLAEAKW-----YTQALNLLGDEPQVWLSTTWNANPLGNAASK 1908
QY 1019 WTLKLHFQEVILGASMRALYIVDPRAWIYMSLYNKSPLSKNYPEHLAKLKFIEGCK 1078
Db 1909 TTOQVR--QQVL--TQLRLSRVKT-----LLGTANSLTAL--FLPQENSKL-----K 1951
QY 1079 G-----KCNLSGVAFEYEPRLKELSKSNVSLLSHLWLANTAAALRINTDLLPT 1130
Db 1952 GYVRTLAQRMFNRLHNSIDGQPLSLPYAKPADPKALLUSAAYSASQGG----- 2001
QY 1131 SYQLVKFEDIVH-FPQ 1145
Db 2002 --DLPKAPLIHFPQ 2015

RESULT 5
US-08-851-567B-53
; Sequence 53, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitiari
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow AgroSciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis

STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.93804
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1844 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-567B-53

Query Match 2.1%; Score 135; DB 4; Length 1844;
Best Local Similarity 17.6%; Pred. No. 0.0053;
Matches 186; Conservative 150; Mismatches 365; Indels 354; Gaps 48;

QY 29 FSTFEESVSNYSEWA-----VFTDDIDQFKTKQVQDFRPNQKLKSLMLPSLYFD 78
Db 916 FTDWERYNKRYSTWAGVSELVYPENYVDPTRIGQTKMD-ALLQSIQSQLNADTVED 974
QY 79 AGEIOAMRQKSRASHLHLFRAIRSAVTV-----MLSNPTYLPPPKGA-----D 122
Db 975 AFKTY-LTSFEQVANLKVISAYHDNVNVDQGLTYFIGIDQAAPGTYVRSVDHSCENK 1033
QY 123 FAA----KWNEIYGNLPPALYCLLCPEDKVAPEFVLEVMYDMVGYKDWLVENAPGDEV 178
Db 1034 FAANAWGWNKI-----TCAVNFWKIIIRVV--YMSRL--YLLWLEQQSKSD- 1078
QY 179 PIGHSLTGFAFADFPLYNLNHRQKYLEKIW-----VITEEMEYYSKVRSWGKQLLH 232
Db 1079 -----DGKTTIYQ--YNLKLAIH--YDGSWNTPTFTDTEKVKNTSTSDAAESL-- 1125
QY 233 NHQATNMIALTLGALVTGVDKSKANIWKQAVVDVWMEKTMELLNHIVDGSLDEGVAGSY 292
Db 1126 -----GLYCTGYQG-----EDTLVVMFYMOSS-----YSSY 1152
QY 293 T--AKSVTQVYVLAQRHFNINNLID--NNWLKQHFVYATILLPGFQRTVGIADSNYNWF 347
Db 1153 TDNNAVPTGLYIFADMSSDNMTNQAATYWNNSY-----PQFDIVMADPDS---- 1199
QY 348 YGPESQLVFLDKFILKNGAGNWLAAQOIRKIRPKQGMVPSVTAQRWSTLHTIYIWDPOL 407
Db 1200 -----NKKVITRRVNNRYABDYE-----IPSSV-----TSNSNYSMGDHSLT 1236


```

QY 408 PGPADYGTAKIHTFPNKGVVYVYAGLPN-----TQNTFVS-FKSGKLGRAV 455
Db 1237 -----MLYGSVPNITFESAEDLRLSTNMAISIIHNGVAGTRRI 1276
QY 456 YDIVHFPYGMIDGWRSFNPGHEPDQNSFTFAPNGQVFVSEALYKPLSHLNNVLVAP 515
Db 1277 QCNLMKQVSLGDKFIYDSSFD--DANRNLVLP---LF-----KFGKDNSSDDISICIYNE 1327
QY 516 SPSSQCHNPGEGQGECAQWLKWTGE-----EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
Db 1328 NPSSSE-DKKWYFSSKDDNKTDADYNGGTCIDAGTSNKDFYNNQIEIVISVTGGYWSYK 1386
QY 572 SAMRLKSVYRALLLNQTLVVDHIERQEDSPNSVSAFFHNLDIDFKIPIPKFMRNRYN 631
Db 1387 IS-----NPNIN-----NTGIDSARV--KVTVKAG 1409
QY 632 GAMMDVNDAYKMFWDHNGSPMASIOEAAEFKKRWTFQVNVTFQWEST--ITRIA 689
Db 1410 G-----DDQIFADNSTVVPQ-----QPAPSPFEMIYQFNLTIDCKNLNFIDNOA 1455
QY 690 YVFGPYINVSCHPFISSNPGLOISLNVNNTVHVSIVTDYHNL-----KTRNLYL 741
Db 1456 HIEIDFTATQDGRFLGAETFIIPVTKVLGTENVIALYSNNGVQVMOYGAYKTRNLTL 1515
QY 742 GFGGFASVADQ-----QITRFLGLTQAIKVPVRHDIRIIPFGFKFNIAVGLI 789
Db 1516 PAQOLVSRANGIDAVLSMETONIQEQLGAGTVVQLVDKYDSIHGTNKSFAIEYVDI 1575
QY 790 LCISLVILTQWRFYLSFRKLMRWILVIALWFIELL-----DVMSTCSQPICAKW---- 841
Db 1576 FKENDSFVIYGELSETSTQTVVK-----VFLSYFIEATGNKNHLW-----VRAKYQKET 1624
QY 842 -----TRTEAE-----GSKKSLSEGHMDLPDVVITSLPGSGA--- 875
Db 1625 TDKILFRTDEKPHGFWLSDDKHTFSGLSAQALKNDSEMPDF-----SGANAL 1674
QY 876 -----EILKQFNSPD--FLYIRVPTAYIDIPELEIDSFVDCAEWKV 918
Db 1675 YFWELFYTPMMAHRLQEQNFDAANHWFFYVWSPSGYI-----VDGKIAIYHNW 1726
QY 919 SDIRSGHFRLLRWLQSLVQDTKLHLQNIHLHFN 953
Db 1727 -----RPLEEDTSNAQQLDSTDTPD 1746

RESULT 6
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-263-023-2

Query Match 2.1%; Score 134; DB 3; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.00054;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 1034 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 1085
Db 278 LKVIHLVRDPR--VASSRIRSRHGLIRSLQVVRSDPRHRMPFLEAAGHLGAKKEGM 336
QY 1086 GYAFEPYELRKLKSKSNVSLSHLWLAANTAAALRINTDLLPTSQVLYKVFEDIVHPQ 1145
Db 337 GGPADYHAL-----GAMEVICNSMAKTLOTALQ--PPDWLQGHVLYVRYEDLVGDPV 386
QY 1146 KTERIFAFIIGIPLSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNWKNQLP 1197
Db 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAARWALT 441
QY 1198 RDEIKLIENICWTLMDRLGYPK 1219
Db 442 FQIKQVEEFCYQPMVILGYER 463

RESULT 7
US-09-471-867-2
; Sequence 2, Application US/09471867
; Patent No. 6455289
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/471,867
; CURRENT FILING DATE: 1999-12-23
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1999-03-05
; EARLIER FILING DATE: 1998-03-05
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-471-867-2

Query Match 2.1%; Score 134; DB 4; Length 483;
Best Local Similarity 26.2%; Pred. No. 0.00054;
Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

QY 1034 MRALYIVRDPRAWIYMLYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLNS 1085
Db 278 LKVIHLVRDPR--VASSRIRSRHGLIRSLQVVRSDPRHRMPFLEAAGHLGAKKEGM 336
QY 1086 GYAFEPYELRKLKSKSNVSLSHLWLAANTAAALRINTDLLPTSQVLYKVFEDIVHPQ 1145
Db 337 GGPADYHAL-----GAMEVICNSMAKTLOTALQ--PPDWLQGHVLYVRYEDLVGDPV 386
QY 1146 KTERIFAFIIGIPLSPA-----SLNQILFATSTNLFYLPYEGEISPTN-----TNWKNQLP 1197
Db 387 KTLRRVYDFVGLLVSPMEQFALNMTSGSGSSKPFV-----VSARNATQAANAARWALT 441
QY 1198 RDEIKLIENICWTLMDRLGYPK 1219
Db 442 FQIKQVEEFCYQPMVILGYER 463

```


RESULT 11

```

US-09-949-016-9449
; Sequence 9449, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9449
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9449

```

RESULT 12

US-09-976-594-168
 ; Sequence 168, Application US/09976594
 ; Patent NO. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 168
 ; LENGTH: 1456
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID NO. 6673549 1674368CD1
 US-09-976-594-168

RESULT 13

```

RESUL1 13
US-09-949-016-7011
; Sequence 7011, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 7011
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7011

Query Match      1.9%; Score 123.5; DB 4; Length 395;
Best Local Similarity 23.3%; Pred. No. 0.0043;
Matches 90; Conservative 53; Mismatches 140; Indels 103; Gaps 20;

Qy 865 VVITSLPGSAGAILKQLFFNSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKV-SDIRS 923
Db 43 VLVLSSWRSGSFVQLFNQHPDVFYLMPEA-----WHVWTTLSQ 82

Qy 924 GHFELLRGWLQSVODTKL-HLQNIHLHEPNRGLAQYFAMNKKKFKFRRESLPEQRS 982
Db 83 GSAATLHMAVRDLVRSVFLCDMDVFDAYLPWRNLSDLF-----QWASRALCSPPACS 136

Qy 983 QMKGAFORDAEYIRALRRHLVYTPS---ARPVLSLSSGSWTLKLHFFO-EVL-----G 1031
Db 137 ----AFPRGAISSAENCKPICARQSFTLAREACRSYSHVVLEKVRFNQLVPLLSDDPA 192

Qy 1032 ASMRALYIVRPRAWIYMLYNKPSLSYLNKNVPEHLAKLFKIEGG--KGKCNLNSGYAF 1089
Db 193 LNLRIVHLVRDPRVLSR-----EQTAKALARDNGIVLG-----TNGTWV 233

Qy 1090 EYEP-LR--KELSKSKSNVSLLSHLWLANTAALRINTDLLTSYQLVKFEDIVHPPQK 1146
Db 234 EADPGLRVREVCR-----SHVRIAE-AATLK-PPFPLRGRVLRVFEFLAREPLA 282

Qy 1147 TTERIFAFGLGIPSP---ASLNQI-----LFATSTNLFYLPYEGEISPTN-TNV 1191
Db 283 EIRALYAFTGLSLTPQLPEAMHNIHSGSGPGARREAFKTSR-----NALNVSA 332

Qy 1192 WKQNLPRDEIKLIENICWTLMDRLGY 1217
Db 333 WRHALPPAKIRRVQELCAGALQLLGY 358

RESULT 14
US-08-899-514-2
; Sequence 2, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; APPLICANT: FUKUTA, MASAKAZU
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-899-514-2

Query Match      1.8%; Score 120.5; DB 2; Length 479;
Best Local Similarity 16.7%; Pred. No. 0.012;
Matches 71; Conservative 64; Mismatches 152; Indels 137; Gaps 12;

Qy 843 RTEAGSKSLSEGHMDLPDVVITSLPGSAGAILKQLFFNSSDFLYIRVP-----894
Db 118 RKEEPPRPVAVAGPRRH-----VLLMATTRTSGSFVGEFFNQGNIFYLFEPLWHIERTV 172

Qy 895 -----TAYIDIPETELEIDSFVDACEWKVSDIRSGH---FRLLRGWLQSLVQ 938
Db 173 SPEPGANNAAGSALVTRDVLKQLFLCDLYV--LEHFIPLPEDHLTQFMFRGSSRLCE 230

Qy 939 D-----TKLHLQNIHLHEPNRGLAQYFAMNKKKFKFRRESLPEQRSOMKGAFDRDA 992
Db 231 DPVCTPFVKVPEKYCKNRCGPNVTLAAEACRRK-----EHMALKAVRIQL 280

Qy 993 EYIRALRHLVYTPARPVLSLSSGSWTLKLHFFOEVLGASMRALYIVRDPRAWIYSMLY 1052
Db 281 EFLQPL-----AEDPRLDL-----RVQLVRDPRVLAASRMV 312

Qy 1053 NSKPSLSYLNKNVPEHLAKLFKIEGGKCNLNSGYAFVPEPLRKLSKSKSNVSLLSHL 1112
Db 313 -----AFAGKYTKWKWLDDEGQGLREEEVQ 339

Qy 1113 WLANTAALRINTDL-----LPTS YQLVKFEDIVHPPQKTFERFAFLGIPLSPASLN 1165
Db 340 RLRGNCESIRLSAELGLRQPAWLRGRVYMLVRYEDVARGPLQKAREMYPPAGIPLTPQVED 399

Qy 1166 QI-----LFATSTNLFYLPYEGEISPTNTVNWKNLPRDEIKLIENICWTLMD 1213
Db 400 WIKNTQAAHDSGIYSTQKN-----SSEQFEKWRFSMPFKLAQVVOAQCPCGAMR 449

Qy 1214 RLGY 1217
Db 450 LFGY 453

RESULT 15
US-09-949-016-6813
; Sequence 6813, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6813
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6813
```

Query Match 1.8%; Score 119.5; DB 4; Length 390;
Best Local Similarity 21.6%; Pred. No. 0.011;
Matches 86; Conservative 54; Mismatches 163; Indels 95; Gaps 16;

Qy	849	SKKSLSEGHMDLPDVVITSLPGSGAEILLKQLFNSDFLYIRVP-----TA	896
Db	28	SRPGFSPAGGEDRVHVLVLSWSRSGSFLGQLFSQHPDVFYLMPEPAWHVWTTLSQGSAA	87
Qy	897	YIDIPETELEIDSFVDACEWKVSDI---RSGHFRLLRGWLQSLVQDTKLHLQNIHLHEP-	952
Db	88	TLHMAVRDLMRISIFL--CDMDVFDAYMPOSRLNLSAFNWTASRA-----LCSP	134
Qy	953	-----NRGKLQYFAMNKKRPFKRRESLPQRSQMKGAFDRDAEYIRALRRHLVYPS	1007
Db	135	ACSAPPRGTISK-----QDVCKTLCTRQPPFSLAREACRSYSHVVLKEVRFNQLVPL	189
Qy	1008	ARPVLSLSSGSWTCLKHFFQEVLGASMRALYIVDPRAWIYMSLYNSKPSLYSLKNVPEH	1067
Db	190	SDPALNL-----RIVHLVRDPRAVLSR--EAAAGPILARDNGI---225	
Qy	1068	LAKLFKIEGGKGNLNSGYAFYEP---LRKELSKSKSNVSLSHLWLANTAAALRN	1124
Db	226	-----VLGTNGK-----VEADPHRLIREVCR-----SHVRIAE-AATLK-P	261
Qy	1125	TDLLPTSQVLYKFEDIVHFPQKTTTERIPAFGLIPSPASLNQILFATSTNLFYLPYEG-E	1183
Db	262	PPFLRGRLVRFEDLAREPLAEIRALYAFGTGLTPQLEAWIHNIHSGSIGKPIEAFH	321
Qy	1184	ISPTN-----TNWQNLPDRDEIKLIENICWTLMDRLG	1217
Db	322	TSSRNARNVSQMRHALPFTKILRVQEVCGALQLLG	359

Search completed: June 23, 2005, 08:52:18
Job time : 50.0594 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:49:14 ; Search time 154.482 Seconds
(without alignments)

3041.886 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWTMALMFTGHLH.....LIENICWTLMDRLGYPKFM 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1717557 seqs, 384547976 residues

Total number of hits satisfying chosen parameters: 1717557

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5532	100.0	1222	16	US-10-697-828-15
2	6509	99.6	1222	16	US-10-475-446-4
3	6465	99.0	1207	17	US-10-479-472A-2
4	5899.5	90.3	1207	16	US-10-697-828-17
5	5899.5	90.3	1207	17	US-10-479-472A-4
6	3147	48.2	596	16	US-10-697-828-9
7	568	8.7	125	11	US-09-984-429-71
8	568	8.7	126	14	US-10-150-111-71
9	147	2.3	411	14	US-10-021-660-128
10	147	2.3	411	15	US-10-211-462-97
11	147	2.3	411	16	US-10-408-765A-395
12	147	2.3	411	16	US-10-408-765A-395

12	147	2.3	411	16	US-10-723-860-1544
13	145	2.2	2504	9	US-09-817-514A-8
14	145	2.2	2504	15	US-10-262-794A-12
15	145	2.2	2504	16	US-10-754-115-59
16	135	2.1	1844	15	US-10-262-794A-53
17	134	2.1	483	14	US-10-212-933-2
18	133	2.0	484	14	US-10-212-933-4
19	133	2.0	530	16	US-10-723-860-1409
20	133	2.0	531	9	US-09-833-790-255
21	133	2.0	531	16	US-10-755-889-284
22	127	1.9	1456	9	US-09-870-759-95
23	127	1.9	1456	10	US-09-751-708A-95
24	127	1.9	1456	15	US-10-295-027-1217
25	127	1.9	1456	16	US-10-428-817A-91
26	127	1.9	1456	16	US-10-723-860-673
27	127	1.9	1456	17	US-10-937-758A-72
28	126	1.9	2322	17	US-10-732-923-13997
29	123.5	1.9	395	9	US-09-927-602-2
30	123.5	1.9	395	15	US-10-258-080-5
31	123.5	1.9	395	16	US-10-648-593-159
32	123.5	1.9	395	16	US-10-697-828-13
33	123.5	1.9	479	13	US-10-087-192-126
34	120	1.8	2360	17	US-10-732-923-13960
35	119.5	1.8	390	9	US-09-927-602-4
36	119.5	1.8	390	16	US-10-697-828-8
37	119	1.8	481	13	US-10-087-192-123
38	118.5	1.8	418	9	US-09-927-602-5
39	118	1.8	395	16	US-10-697-828-7
40	118	1.8	395	16	US-10-841-707-8
41	115.5	1.8	386	14	US-10-427-631-11
42	115	1.8	893	13	US-10-014-436-4
43	115	1.8	1938	13	US-10-014-436-2
44	113.5	1.7	1497	16	US-10-437-963-138435
45	113	1.7	395	9	US-09-927-602-3

ALIGNMENTS

RESULT 1
US-10-697-828-15
; Sequence 15, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT FILING DATE: 2003-10-29
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/144,694
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-15

Query Match	100.0%	Score	6532	DB	16	Length	1222
Best Local Similarity	100.0%	Pred. No.	0				
Matches	1222	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	MPKGGAPPWTMALMFTGHLHFLALLMFAFSTFEESVSNYSEWAVFTDIDQFKTKQVDF	60				
Db	1	MPKGGAPPWTMALMFTGHLHFLALLMFAFSTFEESVSNYSEWAVFTDIDQFKTKQVDF	60				
Qy	61	RPNOKLKKSMHPSLYFDAGBIQAKRSASHLHLFRAISAVTVMLSNPTYYLPPPKH	120				

Db 61 RPNQKLSMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPKPKH 120
Qy 121 ADFAAKWNEIYGNLPLALYCLLPCKEDKAPFVLEYMDBMVGCKDWLVENAPGDSVPI 180
Db 121 ADFAAKWNEIYGNLPLALYCLLPCKEDKAPFVLEYMDBMVGCKDWLVENAPGDSVPI 180
Qy 181 GHSUTGATAPDFLYNLDDHRRRQKYLEKTWITEEMVEYSKVRSGKQLLHNHOATNMI 240
Db 181 GHSUTGATAPDFLYNLDDHRRRQKYLEKTWITEEMVEYSKVRSGKQLLHNHOATNMI 240
Qy 241 ALLTGALVTGDKGSKANIQAQVVDVMEKTMFLNHIVDGSLDEGVAYGYSYTAQSYQY 300
Db 241 ALLTGALVTGDKGSKANIQAQVVDVMEKTMFLNHIVDGSLDEGVAYGYSYTAQSYQY 300
Qy 301 VFLAQRHFNINLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNL 360
Db 301 VFLAQRHFNINLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNLNNL 360
Qy 361 ILKNGAGNLAQAQIRKRPKDGMPVSTAOBWSLTHTEYIWDYDQPLTPQPPADYGTAKIH 420
Db 361 ILKNGAGNLAQAQIRKRPKDGMPVSTAOBWSLTHTEYIWDYDQPLTPQPPADYGTAKIH 420
Qy 421 TFPNMGVVVYTGAGLPNTQTNTFVSFKSGKLGGRVYDVIHFQPYSWIDGWRSNFPGHEHP 480
Db 421 TFPNMGVVVYTGAGLPNTQTNTFVSFKSGKLGGRVYDVIHFQPYSWIDGWRSNFPGHEHP 480
Qy 481 DONSFTFAPNGQVPSALYGPKLSHLNNVLVFPAPSPSSQCNKPWEGQLGECQWLKWTG 540
Db 481 DONSFTFAPNGQVPSALYGPKLSHLNNVLVFPAPSPSSQCNKPWEGQLGECQWLKWTG 540
Qy 541 EEUVDAAAGEIITASOHGEMVFSGEASVSSAMRLKSVYRALLLNSQTLVVVDHIERQ 600
Db 541 EEUVDAAAGEIITASOHGEMVFSGEASVSSAMRLKSVYRALLLNSQTLVVVDHIERQ 600
Qy 601 EDSPIVSVAFFHNLDIDFKYIPYKFNRYNGAMMDVMDAHYKMFVDDHGNSPMAISQIE 660
Db 601 EDSPIVSVAFFHNLDIDFKYIPYKFNRYNGAMMDVMDAHYKMFVDDHGNSPMAISQIE 660
Qy 661 AEQAABFKRWQVNVNWFQWESITIRIAYVYGYPIYVNSCRFIDSNFGLQISLNVNN 720
Db 661 AEQAABFKRWQVNVNWFQWESITIRIAYVYGYPIYVNSCRFIDSNFGLQISLNVNN 720
Qy 721 TEHVVSIVTDVHNLKTRFNLVGFSGFASVADQGOITRFGLTQAIKVPVRHDIRIIPFPG 780
Db 721 TEHVVSIVTDVHNLKTRFNLVGFSGFASVADQGOITRFGLTQAIKVPVRHDIRIIPFPG 780
Qy 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELLDVNSTCSQPICAK 840
Db 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELLDVNSTCSQPICAK 840
Qy 841 WTRTEABGSKSLSSGHHMDLPDVVITSLPGSABILKQLFNNSSDFLYIRVPTAYIDI 900
Db 841 WTRTEABGSKSLSSGHHMDLPDVVITSLPGSABILKQLFNNSSDFLYIRVPTAYIDI 900
Qy 901 PETELEIDSFDVACEWKVSDIRSGHFRLLRGWLSQVQDTKLHQLNTHLHPNKGKLAQY 960
Db 901 PETELEIDSFDVACEWKVSDIRSGHFRLLRGWLSQVQDTKLHQLNTHLHPNKGKLAQY 960
Qy 961 FANMKDKRFRKRESLPEQSRQMGAFDRDAEYIRALRRLHVLVYPSARPVLISLSSGSWT 1020
Db 961 FANMKDKRFRKRESLPEQSRQMGAFDRDAEYIRALRRLHVLVYPSARPVLISLSSGSWT 1020
Qy 1021 LKLFHFFQEVLGASMRALYIVRDPRAWIYSMLYNSKPSLYSKNVPHEHLAKLFKIEGKGK 1080
Db 1021 LKLFHFFQEVLGASMRALYIVRDPRAWIYSMLYNSKPSLYSKNVPHEHLAKLFKIEGKGK 1080
Qy 1081 CNLNSGVAFEYELURKEISKSKSNVSLSHLWLANTAARLNTDLLPTSQYLKVFEDI 1140
Db 1081 CNLNSGVAFEYELURKEISKSKSNVSLSHLWLANTAARLNTDLLPTSQYLKVFEDI 1140
Qy 1141 VHPFQKTERIFAFGLIGPLSPASLNQILFATSTNLFLPYEGEISPTNTNVWQNLPRDE 1200
Db 1141 VHPFQKTERIFAFGLIGPLSPASLNQILFATSTNLFLPYEGEISPTNTNVWQNLPRDE 1200

Qy 1201 IKLIENICWTMLDRLCYVPKFM 1222
Db 1201 IKLIENICWTMLDRLCYVPKFM 1222
RESULT 2
US-10-475-446-4
; Sequence 4, Application US/10475446
; Publication No. US20040198651A1
; GENERAL INFORMATION:
; APPLICANT: KLAMMER, Aaron A.; HAFALIA, April J.A.
; APPLICANT: DUGGAN, Brendon M.; WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.; TRIBULEY, Catherine M.
; APPLICANT: ARVIZU, Chandra S.; HONCHELL, Cynthia D.
; APPLICANT: NGUYEN, Dannel B.; KALLICK, Deborah A.
; APPLICANT: YUE, Henry; AU-YOUNG, Janice K.
; APPLICANT: RAMKUMAR, Jayalaxmi; LI, Joana X.
; APPLICANT: THANGAVELU, Kavitha; GIETZEN, Kimberly J.
; APPLICANT: DING, Li; BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.
; APPLICANT: MASON, Patricia M.; LAL, Preeti G.
; APPLICANT: GRAU, Richard C.; REDDY, Roopa M.
; APPLICANT: BECHA, Shanya D.; KAREHT, Stephanie K.
; APPLICANT: RICHARDSON, Thomas W.; TRAN, Uyen K.
; APPLICANT: ELLIOTT, Vicki S.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; APPLICANT: XU, Yuming
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PF-0949 USN
; CURRENT APPLICATION NUMBER: US/10/475,446
; CURRENT FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: PCT/US02/12464
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,207
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/287,114
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/288,640
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: US 60/290,516
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/292,184
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/343,553
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/357,002
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: US 60/358,279
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: US 60/366,041
; PRIOR FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 3441255CD1
US-10-475-446-4

Query Match 99.6%; Score 6509; DB 16; Length 1222;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1219; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MPKGGAPPITMALMFTGHLFLALLMFASTFEESVSNYSEWAVFTDDIDQFKTKQVDF 60
Db 1 MPKGGAPPITMALMFTGHLFLALLMFASTFEESVSNYSEWAVFTDDIDQFKTKQVDF 60
Qy 61 RPNQKLSMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPKPKH 120

Db 61 RPNQKKSLHPSLSYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
 Qy 121 ADPAKKNIEYGNLPLALYCLLCPDKVAFEVLEVMRVMGYKDWLVENAPGDEVPI 180
 Db 121 ADPAKKNIEYGNLPLALYCLLCPDKVAFEVLEVMRVMGYKDWLVENAPGDEVPI 180
 Qy 181 GHSITGPATAPDFLYNLLDNHRRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMI 240
 Db 181 VHSITGPATAPDFLYNLLDNHRRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMI 240
 Qy 241 ALLTGALVTGVDGSKANIKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQY 300
 Db 241 ALLTGALVTGVDGSKANIKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQY 300
 Qy 301 VFLAQRHFNINLNNLNNLKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKF 360
 Db 301 VFLAQRHFNINLNNLNNLKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKF 360
 Qy 361 ILKNGAGNWLAAQQIRKHPKDGMPVSTAQKRWSTLHTEYIWDYDQTPQPPADYGTAKIH 420
 Db 361 ILKNGAGNWLAAQQIRKHPKDGMPVSTAQKRWSTLHTEYIWDYDQTPQPPADYGTAKIH 420
 Qy 421 TFPNMGVVTYGAGLPNTQTNTTFVSKGLGGRVAVDVIHFQPSYWDGWRSPNPGHEHP 480
 Db 421 TFPNMGVVTYGAGLPNTQTNTTFVSKGLGGRVAVDVIHFQPSYWDGWRSPNPGHEHP 480
 Qy 481 DONSFTFAPNGQVFSBALYGPKLSHLNNVLVFPAPSSQCNKPWGQGECAQWLKWTG 540
 Db 481 DONSFTFAPNGQVFSBALYGPKLSHLNNVLVFPAPSSQCNKPWGQGECAQWLKWTG 540
 Qy 541 EEVGDAAGEIITASQHGEMFVSGEASVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 Db 541 EEVGDAAGEIITASQHGEMFVSGEASVSSAMRLKSVYRALLLLNSQTLVVVDHIERQ 600
 Qy 601 EDSINSVSAFFHNLDIDFKYIPYKFNRYNGAMVDVDAHYKWFVFDHNGNSPMASIQE 660
 Db 601 EDSINSVSAFFHNLDIDFKYIPYKFNRYNGAMVDVDAHYKWFVFDHNGNSPMASIQE 660
 Qy 661 AEQAAEFKKRWTFQVNTTQMESITITRAYVYGPYINVSCTFIDSNPGLOISLVANN 720
 Db 661 AEQAAEFKKRWTFQVNTTQMEPTITRIAYVYGPYINVSCTFIDSNPGLOISLVANN 720
 Qy 721 TEHVVSIVTDYHNLKTRFNLYLGGFGFASVADQGI TRFGLGTQAIKVPVRHDIRIIFPGF 780
 Db 721 TEHVVSIVTDYHNLKTRFNLYLGGFGFASVADQGI TRFGLGTQAIKVPVRHDIRIIFPGF 780
 Qy 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELDVWSTCSQPICAK 840
 Db 781 KFNIAVGLILCISLVILTQWRFYLSFRKLMRWILILVIALWFIELDVWSTCSQPICAK 840
 Qy 841 WTRTEAGSKSLSEGGHMDLPDVVITSLPGSABILKQFPNNSDPLXIRVPTAYIDI 900
 Db 841 WTRTEAGSKSLSEGGHMDLPDVVITSLPGSABILKQFPNNSDPLXIRVPTAYIDI 900
 Qy 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHQNIIHLHPNKGKLAQY 960
 Db 901 PETELEIDSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHQNIIHLHPNKGKLAQY 960
 Qy 961 FANMKDKRKPRESLPEQSRQKMGAPDRDAEYIRALRRHLVYYPSPARVLSLSSGWT 1020
 Db 961 FANMKDKRKPRESLPEQSRQKMGAPDRDAEYIRALRRHLVYYPSPARVLSLSSGWT 1020
 Qy 1021 LKLHFFQEVLGASMRALYIVDPDRAWIYSMLYNKPSLYSLKNVPEHLAKLFKEGGK 1080
 Db 1021 LKLHFFQEVLGASMRALYIVDPDRAWIYSMLYNKPSLYSLKNVPEHLAKLFKEGGK 1080
 Qy 1081 CNLNSGYAFEPLEPKLSKSNVSLLSHLWLANATAAALRINTDLLPTSYQLVKFEDI 1140
 Db 1081 CNLNSGYAFEPLEPKLSKSNVSLLSHLWLANATAAALRINTDLLPTSYQLVKFEDI 1140
 Qy 1141 VHFQKTTTERIFAFIGIPLSPASLNQILFATSTNLFYLPYGEISPTNTNWKQLPRDE 1200
 Db 1141 VHFQKTTTERIFAFIGIPLSPASLNQILFATSTNLFYLPYGEISPTNTNWKQLPRDE 1200

Qy 1201 IKLIENICWTLMDRLGYPKEMD 1222
 Db 1201 IKLIENICWTLMDRLGYPKEMD 1222

RESULT 3

US-10-479-472A-2
 ; Sequence 2, Application US/10479472A
 ; Publication No. US20050118581A1
 ; GENERAL INFORMATION:
 ; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
 ; APPLICANT: VAN BROECKHOVEN, CHRISTINE
 ; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
 ; TITLE OF INVENTION: BIPOLAR DISORDER
 ; FILE REFERENCE: JAB-1711
 ; CURRENT APPLICATION NUMBER: US/10/479,472A
 ; CURRENT FILING DATE: 2003-12-01
 ; PRIOR APPLICATION NUMBER: PCT/EP02/06316
 ; PRIOR FILING DATE: 2002-06-06
 ; PRIOR APPLICATION NUMBER: EP 01202214.1
 ; PRIOR FILING DATE: 2001-06-11
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: Patencin ver. 3.2
 ; SEQ ID NO 2
 ; LENGTH: 1212
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-479-472A-2

Query Match 99.0%; Score 6465; DB 17; Length 1212;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1211; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 MALMFTGHLFLALLMPAFSTFESVNSYSEWAVFTDDIDQFKTKQVDFPNOKLKKSM 70
 Db 1 MALMFTGHLFLALLMPAFSTFESVNSYSEWAVFTDDIDQFKTKQVDFPNOKLKKSM 60
 Qy 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADFAAKWNEI 130
 Db 61 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADFAAKWNEI 120
 Qy 131 YGNLPLALYCLLCPDKVAFEVLEVMRVMGYKDWLVENAPGDEVPIGHSITGPATA 190
 Db 121 YGNLPLALYCLLCPDKVAFEVLEVMRVMGYKDWLVENAPGDEVPIGHSITGPATA 180
 Qy 191 FDFLYNLLDNHRRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMIALLTGALVTG 250
 Db 181 FDFLYNLLDNHRRQKYLEKIWIITEEMYESKVRSGKQLLHNHQAENMIALLTGALVTG 240
 Qy 251 VDKGSKANIKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQYVFLAQRHFN 310
 Db 241 VDKGSKANIKQAVVDMVMEKTMFLLNHI VDGSLDEGVAYGSYTAQSVTQYVFLAQRHFN 300
 Qy 311 NNLDNNWLKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKFLLKNGAGNWL 370
 Db 301 NNLDNNWLKHFWFYATLLPGFQRTVGIADSNYNWFYGPESQLVFLDKFLLKNGAGNWL 360
 Qy 371 AQQIRKHPKDGMPVSTAQKRWSTLHTEYIWDYDQTPQPPADYGTAKIHFPNMGVVTY 430
 Db 361 AQQIRKHPKDGMPVSTAQKRWSTLHTEYIWDYDQTPQPPADYGTAKIHFPNMGVVTY 420
 Qy 431 GAGLPNTQTNTTFVSKGLGGRVAVDVIHFQPSYWDGWRSPNPGHEHPDQNSFTFAPN 490
 Db 421 GAGLPNTQTNTTFVSKGLGGRVAVDVIHFQPSYWDGWRSPNPGHEHPDQNSFTFAPN 480
 Qy 491 GOVFSVSEALYGPKLSHLNNVLVFPAPSSQCNKPWGQGECAQWLKWTGEEVDAAGEI 550
 Db 481 GOVFSVSEALYGPKLSHLNNVLVFPAPSSQCNKPWGQGECAQWLKWTGEEVDAAGEI 540
 Qy 551 ITASQHGEMFVSGEASVSSAMRLKSVYRALLLLNSQTLVVVDHIERQSDPSINSVSA 610
 Db 541 ITASQHGEMFVSGEASVSSAMRLKSVYRALLLLNSQTLVVVDHIERQSDPSINSVSA 600

611 FFHNLIDDFKYIPYKFNRYNGAMMDVMDAHYKMFDFDHGNSPMASIQEAEQAAEFKKR 670
601 FFHNLIDDFKYIPYKFNRYNGAMMDVMDAHYKMFDFDHGNSPMASIQEAEQAAEFKKR 660
671 WTQFVNVVTFQMESTITRIAYVFGPYINVSRCRFIDSSNFGLOISLVNNVTEHVVISVTD 730
661 WTQFVNVVTFQMESTITRIAYVFGPYINVSRCRFIDSSNFGLOISLVNNVTEHVVISVTD 720
731 YHNLKTRFNVLGGGFPASVADQGOITRFGLGTOAIVKPVVRHDIRIIPPGFKFNIAVGLIL 790
721 YHNLKTRFNVLGGGFPASVADQGOITRFGLGTOAIVKPVVRHDIRIIPPGFKFNIAVGLIL 780
791 CISLVILTFQWRFLVSRKLRWILILVIALWFLLELLDVWSTSCQPICAKWTTEAGSK 850
781 CISLVILTFQWRFLVSRKLRWILILVIALWFLLELLDVWSTSCQPICAKWTTEAGSK 840
851 KSLSEGHMDLDPVVTITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELBIDSF 910
841 KSLSEGHMDLDPVVTITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELBIDSF 900
911 VDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKK 970
901 VDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKK 960
971 FKRESLPEORSOMKGAFFDRAEYIRALRRHLVYPSARPVLISISSGSWTLKLFHFQEV 1030
961 FKRESLPEORSOMKGAFFDRAEYIRALRRHLVYPSARPVLISISSGSWTLKLFHFQEV 1020
1031 GASMRALYIVRDPRAWIYSLMYNSKPSLSYSLKNVPEHLAKLFIEGKGKCNLSGYAFE 1090
1021 GASMRALYIVRDPRAWIYSLMYNSKPSLSYSLKNVPEHLAKLFIEGKGKCNLSGYAFE 1080
1091 YEPLRKELSKSNVNSLLSHLWANTAAALRINTDLLPTSQYLVKPEDIVHPQKTTTER 1150
1081 YEPLRKELSKSNVNSLLSHLWANTAAALRINTDLLPTSQYLVKPEDIVHPQKTTTER 1140
1151 IFAPLGIPLSPASINQILFATSNLFLVYPVEGISPTNTNVKONLPROEIKLIENICWT 1210
1141 IFAPLGIPLSPASINQILFATSNLFLVYPVEGISPTNTNVKONLPROEIKLIENICWT 1200
1211 LMDRLGYPKFMD 1222
1201 LMDRLGYPKFMD 1212

RESULT 4

US-10-697-828-17
; Sequence 17, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; PRIOR FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: mouse
US-10-697-828-17

Query Match

Best Local Similarity 90.34; Score 5899.5; DB 16; Length 1207;

Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

11 MALMFTGHLFLALLMEAFSTPEESVSNYSEWAVFTDDIDQFKTKOKVQDFRPNOKLKSM 70
1 MAFMFTGHLFLUTLMCMCSFSTCEESVSNYSEWAVFTDDIQWLKSKQIQDFKUNRR----- 55
71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTTYLPPPKHADFAAKWNEI 130
56 LHPNLVFDAGDIQTLKQKSRSTSLHLFRAIKSAVTIMLSNPYSYLPPPKHASEFAAKWNEI 115
131 YGNLPPALYCLLPEDKVAPEFVLEYNMDRWGVKMDLVENAPGDEVPIGHSLTGCFATA 190
116 YGNLPPALYCLLPEDKVAPEFVMEYMDRWVSKMDLVENAPGDEVPGVHSLTGCFATA 175
191 FDFLNNLDNHRQKYLEKIWIITEEMVEYSKVRSGKOLLHNHOATNMIALLTGALVTG 250
176 FDFLNNLDNHRQKYLEKIWIITEEMVEYSKIRSGKOLLHNHOATNMIALLTGALVTG 235
251 VDKGSKANIWKQAVVDVMEKTMFLNHLIVDGLDEGVAYGYSYAKSVTYQYVFLAQRHFI 310
236 VDKGSKANIWKQAVVDVMEKTMFLNHLIVDGLDEGVAYGYSYAKSVTYQYVFLAQRHFI 295
311 NNLNLDNHLKQHPFYATLLPGQRTVGIADSNYNWFYGPESQLVFLDKFILQNGAGNWL 370
296 NNLNLDNHLKQHPFYATLLPGQRTVGIADSNYNWFYGPESQLVFLDKFILQNGAGNWL 355
371 AQQIRKHPKQGPMPVSTAQRWSTLHTEYIWDYDQDLPQPPADYGTAKTHTFPNKGVTY 430
356 AQQIRKHPKQGPMPVSTAQRWSTLHTEYIWDYDQDLPQPPADYGTAKTHTFPNKGVTY 415
431 GAGLNTQNTNTFVSPKSGKLGGRAVYDIVHFQYPSWIDGWRSPNFGHEHPDQNSTFPAPN 490
416 GAGLNTQNTNTFVSPKSGKLGGRAVYDIVHFQYPSWIDGWRSPNFGHEHPDQNSTFPAPN 475
491 GQVFVSEALYGPKLHNLNVLVPAPSPSQCNKPWEGQGECAQWLKWTGEVSGAAGEI 550
476 GQVFVSEALYGPKLHNLNVLVPAPSPSQCNKPWEGQGECAQWLKWTGEVSGAAGEV 535
551 ITASQHGEMVFGSGRAVSAYSSAMELKSIVYRALLLNSQTLVVDHIERQEDSPINSVA 610
536 ITAAQHGORMFVSGRAVSAYSSAMELKSIVYRALLLNSQTLVVDHIERQETSPINSVA 595
611 FPHNLIDDFKYIPYKFNRYNGAMMDVMDAHYKMFDFDHGNSPMASIQEAEQAAEFKKR 670
596 FPHNLIDDFKYIPYKFNRYNGAMMDVMDAHYKMFDFDHGNSPMASIQEAEQAAEFKKR 655
671 WTQFVNVVTFQMESTITRIAYVFGPYINVSRCRFIDSSNFGLOISLVNNVTEHVVISVTD 730
656 WTQFVNVVTFQMESTITRIAYVFGPYINVSRCRFIDSSNFGLOISLVNNVTEHVVISVTD 715
731 YHNLKTRFNVLGGGFPASVADQGOITRFGLGTOAIVKPVVRHDIRIIPPGFKFNIAVGLIL 790
716 YONLKSFRSYLGGGFPASVADQGOITRFGLGTOAIVKPVVRHDIRIIPPGFKFNIAVGLIL 775
791 CISLVILTFQWRFLVSRKLRWILILVIALWFLLELLDVWSTSCQPICAKWTTEAGSK 850
776 CISLVILTFQWRFLVSRKLRWILILVIALWFLLELLDVWSTSCQPICAKWTTEAGSK 835
851 KSLSEGHMDLDPVVTITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELBIDSF 910
836 KYMISEGHHVDLPNVITITSLPGSGAEILKQLFNSSDFLYIRVPTAYIDIPETELBIDSF 895
911 VDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKK 970
896 VDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHEPNRGKLAQYFAMNKDKK 955
971 FKRESLPEORSOMKGAFFDRAEYIRALRRHLVYPSARPVLISISSGSWTLKLFHFQEV 1030
956 LKRESLQDQRSRIKGPFDRAEYIRALRRHLVYPSARPVLISISSGSWTLKLFHFQEV 1015
1031 GASMRALYIVRDPRAWIYSLMYNSKPSLSYSLKNVPEHLAKLFIEGKGKCNLSGYAFE 1090
1016 GTSMEALYIVRDPRAWIYSLMYNSKPSLSYSLKNVPEHLAKLFIEGKGKCNLSGYAFE 1075

```
Qy 1091 YELRLKELSKSNVALLSHLWANTAAALINTDLPSTSYQLVKPEDIHVPQKTTT 1150
Db 1076 YESLKELETSQSNALSHLWANTAAALINTDLPSTSYQLVKPEDIHVPQKTTT 1135
Qy 1151 IFAPLGIPLSPASINQILFATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
Db 1136 IFAPLGIPLSPASINQILFATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
Qy 1211 LMDRLGYPKFMD 1222
Db 1196 LMDHLGYPKFMD 1207

RESULT 5
US-10-479-472A-4
; Sequence 4, Application US/10479472A
; Publication No. US2005018581A1
; GENERAL INFORMATION:
; APPLICANT: DEL-FAVERO, JURGEN PETER LODE
; APPLICANT: VAN BROECKHOVEN, CHRISTINE
; TITLE OF INVENTION: NOVEL BRAIN EXPRESSED GENE AND PROTEIN ASSOCIATED WITH
; TITLE OF INVENTION: BIPOLAR DISORDER
; FILE REFERENCE: JAB-1711
; CURRENT APPLICATION NUMBER: US/10/479,472A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: PCT/EP02/06316
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: EP 01202214.1
; PRIOR FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 4
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: Mus sp.
US-10-479-472A-4
Query Match 90.3%; Score 5899.5; DB 17; Length 1207;
Best Local Similarity 89.4%; Pred. No. 0;
Matches 1084; Conservative 68; Mismatches 55; Indels 5; Gaps 1;

Qy 11 MALMFTGHLFLALLMFAFSTFEESVSNYSSEAWFTDDIDQFKTKQVQDPRNQKLKSM 70
Db 1 MAPMTEHLLFLTLMMCSFSTCEESVSNYSSEAWFTDDIQWLKSKQIQDPKLNRR---- 55
Qy 71 LHPSLYFDAGEIQAMROKRSRASHLHLFRAIRSAVTMLSNPTYYLPPLPKHADPAKNEI 130
Db 56 LHPNLVFDAGDIQTLKQKRSRTSHLHIFRAIKSAVTIMLSNPSYYLPPLPKHAEFAAKNEI 115
Qy 131 YGNLPLALYCLLCPEDKVAFEFLVYMDRMVGYKDWLVENAPGDEVPICHSITGPATA 190
Db 116 YGNLPLALYCLLCPEDKVAFEFLVYMDRMVGYKDWLVENAPGDEVPVGSHTGPATA 175
Qy 191 FDFLYNLDDHRRQKYLEIKIWIITEEMVEYSKVRSGKQLLHNNHQAATNMIALLTGALVTG 250
Db 176 FDFLYNLGNQKQKYLEIKIWIITEEMVEYSKVRSGKQLLHNNHQAATNMIALLTGALVTG 235
Qy 251 VDKGSKANIWKQAVDVNKEKTMFLNLNHIVDGSLDEGVAYGSAKSYQYVFLAQRHFI 310
Db 236 VDKGSKANIWKQAVDVNKEKTMFLNLNHIVDGSLDEGVAYGSAKSYQYVFLAQRHFI 295
Qy 311 NLDNNMLKMHFWFYATLLPGFORTVGIADSNYNWFGPESQVFLDKFKLKNAGNWL 370
Db 296 NFDNNMLKMHFWFYATLLPGFORTVGIADSNYNWFGPESQVFLDKFKLKNAGNWL 355
Qy 371 AQQIRKHPKDGMPVPSTAQRWSTLHTEYIWDYDLPQTPQPPADYCTAKIHTFPNKGVVY 430
Db 356 AQQIRKHPKDGMPVPSTAQRWSTLHTEYIWDYDLPQTPQPPADYCTAKIHTFPNKGVVY 415
Qy 431 GAGLPNTQNTVTSFKSKLGGRAVYDIIHVPQYSWIDGWSFNPGEHHPDQNSFTFAPN 490
Db 416 GGGLPNTQNTVTSFKSKLGGRAVYDIIHVPQYSWIDGWSFNPGEHHPDQNSFTFAPN 475
```

```
Qy 491 GQVVFSEALYGPKLSHLNNVLPFAPSPSSQCNKPNWEGQLGECAQWLKWTGEEVGDAAGEI 550
Db 476 GQVVFSEALYGPKLSHLNNVLPFAPSPSSQCNKPNWEGQLGECAQWLKWTGEEVGDAAGEV 535
Qy 551 ITASQHGEMVFGSBAVSAVSSAMRLKSVYRAILLNSQTLVVVDHIEROEDSPINSVSA 610
Db 536 ITAAQHGDRMFVSGEAVSAVSSAMRLKSVYRAILLNSQTLVVVDHIEROETSPINSVSA 595
Qy 611 FFHNLDDIDFKYIPYKPMNRNGAMDMVDAHYKMFDFHHGNSPMASIQBAEQAAEFKKR 670
Db 596 FFHNLDDIDFKYIPYKPMNRNGAMDMVDAHYKMFDFHHGNSPMVANIQBAEQAAEFKKR 655
Qy 671 WFOFVNVTFOMESTITRIAYVFGYPIYVSSCRPIDSSNPGLOISLNNVTEHVVSVITD 730
Db 656 WFOFVNVTFHMESTITRIAYVFGYPIYVSSCRPIDSSNPGLOISLNNVTEHVSVMVD 715
Qy 731 YHNLKTRFNLGFGGFPASVADQGITRFGIGTQAIKVPVRHRIIPFPKFNIAVGLIL 790
Db 716 YQNLKGRFSYLGFGGFPASVANQGITRFGIGTQAIKVPVRHRIIPFPKFNIAVGLIL 775
Qy 791 CISLVILTFQWRFYLPFRKLMRWILILVIALWFIELLDVMTSCSQPICAKWTRTEAESK 850
Db 776 CISLVILTFQWRFYLPFRKLMRWILILVIALWFIELLDVMTSCSQPICAKWTRTEAKNE 835
Qy 851 KSLSSGHHMDLPDVITSLPGSGAEILKQLPFNSSDFLIRVPTAVIDIPETELEDISF 910
Db 836 KVMISEGHVLDLPNVITSLPGSGAEILKQLPFNSSDFLIRVPTAVIDIPETELEDISF 895
Qy 911 VDACEKVSDIRSGHFLHLLGWSLQVODTKLHLNTHLHEPNRGLAQPAMNKDKK 970
Db 896 VDACEKVSDIRSGHFLHLLGWSLQVODTKLHLNTHLHEPNRGLAQPAMNKDKK 955
Qy 971 FKRESLPEQORSQMGAFDRAEYIRALRRLHLYVYPSARPVLSLSSGSWTLKLFHFFQEV 1030
Db 956 LKRESLQDORSRIKGFDDAEYIRALRRLHLYVYPSARPVLSLSSGSWTLKLFHFFQEV 1015
Qy 1031 GASMRALYIVRDPRAWIYSLYKSPSLYSLKNVPEHLAKFKITEGKGKCNLSGVAFE 1090
Db 1016 GTSMRALYIVRDPRAWIYSLYKSPSLYSLKNVPEHLAKFKITEGKGKCNLSGVAFE 1075
Qy 1091 YELRLKELSKSNVALLSHLWANTAAALINTDLPSTSYQLVKPEDIHVPQKTTT 1150
Db 1076 YESLKELETSQSNALSHLWANTAAALINTDLPSTSYQLVKPEDIHVPQKTTT 1135
Qy 1151 IFAPLGIPLSPASINQILFATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1210
Db 1136 IFAPLGIPLSPASINQILFATSTNLFYLPYEGEISPTNTNWKONLPRDEIKLIENICWT 1195
Qy 1211 LMDRLGYPKFMD 1222
Db 1196 LMDHLGYPKFMD 1207

RESULT 6
US-10-697-828-9
; Sequence 9, Application US/10697828
; Publication No. US20040185546A1
; GENERAL INFORMATION:
; APPLICANT: Rosen, Steven
; APPLICANT: Lee, Jin Kyu
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: Novel Glycosyl Sulfotransferases GST-4alpha, GST-4beta, & GST-6
; FILE REFERENCE: UCAL-138DIV
; CURRENT APPLICATION NUMBER: US/10/697,828
; CURRENT FILING DATE: 2003-10-29
; PRIOR APPLICATION NUMBER: 09/593,828
; PRIOR FILING DATE: 2000-06-13
; PRIOR APPLICATION NUMBER: 60/144,694
; PRIOR FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 596
```

```
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-697-828-9

Query Match
Best Local Similarity 48.2%; Score 3147; DB 16; Length 596;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 MNRVNGAMVDVDAHYKQFDFHGHGSPMASIOEASQAEFKKRWTFQVNVTFQMESTIT 686
Db 1 MNRVNGAMVDVDAHYKQFDFHGHGSPMASIOEASQAEFKKRWTFQVNVTFQMESTIT 60

QY 687 RIAYVFYGPINVSRCFIDSSNPGQLISLVNNTTHVSVIVDYHNLKTRFNLYLFGGFG 746
Db 61 RIAYVFYGPINVSRCFIDSSNPGQLISLVNNTTHVSVIVDYHNLKTRFNLYLFGGFG 120

QY 747 ASVADQOQITRFGLTQAIKVPVRHDIRIIPFGFKFNIAVGLILCIISLVILTQPMRFYLS 806
Db 121 ASVADQOQITRFGLTQAIKVPVRHDIRIIPFGFKFNIAVGLILCIISLVILTQPMRFYLS 180

QY 807 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVV 866
Db 181 FRKLMRWILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVV 240

QY 867 ITSPLSGABILKQLFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 926
Db 241 ITSPLSGABILKQLFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVSDIRSGHF 300

QY 927 RLRLGWLQSLVQDTKLHQLNIHLEPNRGKLAQYFAMNKDKKGFKRRESLPLPQRSQMGK 986
Db 301 RLRLGWLQSLVQDTKLHQLNIHLEPNRGKLAQYFAMNKDKKGFKRRESLPLPQRSQMGK 360

QY 987 AFORDAEYIRALRRHLVYVYAPVLSLSSGSKWTLKHFEQVGLGASMRALYIVRDPRAW 1046
Db 361 AFORDAEYIRALRRHLVYVYAPVLSLSSGSKWTLKHFEQVGLGASMRALYIVRDPRAW 420

QY 1047 IYSLMYSKPSLSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEPYPLRKLKSLSKSNV 1106
Db 421 IYSLMYSKPSLSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEPYPLRKLKSLSKSNV 480

QY 1107 SLLSHLWLANTAALRINTDLPSTVQLVKFEDIHFPQKTTTERIFAFGLIPSPASLQ 1166
Db 481 SLLSHLWLANTAALRINTDLPSTVQLVKFEDIHFPQKTTTERIFAFGLIPSPASLQ 540

QY 1167 ILFATSNLFLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 1222
Db 541 ILFATSNLFLPYEGEISPTNTNVKQNLPRDEIKLIENICWTLMDRLGYPKPM 596

RESULT 7
US-09-984-429-71
; Sequence 71, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; CURRENT FILING DATE: 2001-10-30
; PRIOR FILING DATE: 60/244,591
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-10-150-111-71

Query Match
Best Local Similarity 8.7%; Score 568; DB 14; Length 126;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 811 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVVITSL 870
Db 1 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVVITSL 60

QY 871 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 917
Db 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 8
US-10-150-111-71
; Sequence 71, Application US/10150111
; Publication No. US20030078386A1
; GENERAL INFORMATION:
; APPLICANT: Rubin et al.
; TITLE OF INVENTION: Secreted Protein HPEAD48
; FILE REFERENCE: P2018PDI
; CURRENT APPLICATION NUMBER: US/10/150,111
; CURRENT FILING DATE: 2002-05-20
; PRIOR FILING DATE: 09/288,143
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 71
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (126)
; OTHER INFORMATION: Xaa equals stop translation
US-10-150-111-71

Query Match
Best Local Similarity 100.0%; Pred. No. 9.1e-43;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 811 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVVITSL 870
Db 1 MRWILILVIALWFIELLDVWSTCSQPICAKWTTRTEAGSKKSLSEGHGHHMDLPDVVITSL 60

QY 871 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 917
Db 61 PGSGAEILKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107
```

Db 61 PGSGAELKQLFFNSDFLYIRVPTAYIDIPETELEIDSFVDACEWK 107

RESULT 9

US-10-021-660-128
; Sequence 128, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: NO. US20030152926A1el Methods of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT APPLICATION NUMBER: US/10/021,660
; CURRENT FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-128

Query Match 2.3%; Score 147; DB 14; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 813 WILILVIALWFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 855
Db 5 WKAVLLALASIAIQYTAIRTTAKSFHTCPGLAEGLAERLC-----EESPTAYNLRS 59
Qy 856 EGHMDLPDVVITSLPGSGAELKQLFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db 60 KTH-----ILILATTRSGSFGVQLFNQHLDFVYLFEPL----- 93
Qy 916 WKVSDIRSGHFRLLRGLWQLSLVQDTKLHLQNIHLHFNPNRGKLAQYFAMNKDKKFKRRE 975
Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115
Qy 976 SLPEORSOMKGAFRDAEYI-----RALRRHLVYVPSARPV-----LS 1013
Db 116 MLGASRDLLRSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175
Qy 1014 LSSGSWTCLKHFFQEVILGA-----SMRALYIVRD 1042
Db 176 LEEGDCVYKCGLLNLTVAAEACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVRD 235
Qy 1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYEPRLKELSKS 1102
Db 236 PRGILASRSETFRDY-----RLWRLWYGTGR-----KPNLDVTQL-TTVCSDF 279
Qy 1103 SNAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEIVHFPQKTFERIFAFIGIPL- 1159
Db 280 SNSVSTGLMRPPLKAG-----KYMVLVRYEDLARNPMKTEIIVGFLGIPLD 325
Qy 1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTNVWKNLPRDEIKLIENI 1207
Db 326 SHVARWIONNTRGDPITLGHKHYGTVRN-----SAATAEKWRFLSYDIVAFAPNA 375
Qy 1208 CWTLMORLGY 1217
Db 376 CQOVLAQLGY 385

RESULT 10

US-10-211-462-97

; Sequence 97, Application US/10211462
; Publication No. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 97
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-462-97

Query Match 2.3%; Score 147; DB 15; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

Qy 813 WILILVIALWFIEL-----LDVWSTC-----SOPICAKWTRTEAGSKSLSS 855
Db 5 WKAVLLALASIAIQYTAIRTTAKSFHTCPGLAEGLAERLC-----EESPTAYNLRS 59
Qy 856 EGHMDLPDVVITSLPGSGAELKQLFNSDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db 60 KTH-----ILILATTRSGSFGVQLFNQHLDFVYLFEPL----- 93
Qy 916 WKVSDIRSGHFRLLRGLWQLSLVQDTKLHLQNIHLHFNPNRGKLAQYFAMNKDKKFKRRE 975
Db 94 -----YHVQNTLIPRFTQG-----KSPADRRV 115
Qy 976 SLPEORSOMKGAFRDAEYI-----RALRRHLVYVPSARPV-----LS 1013
Db 116 MLGASRDLLRSLYDCDLYFLENYIKPPPVNHTDRIFRGASRVLCSPVCDPPGPADLV 175
Qy 1014 LSSGSWTCLKHFFQEVILGA-----SMRALYIVRD 1042
Db 176 LEEGDCVYKCGLLNLTVAAEACRERSHVAIKTVRVPEVNDLRALVEDPRLNLKVIQLVRD 235
Qy 1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYEPRLKELSKS 1102
Db 236 PRGILASRSETFRDY-----RLWRLWYGTGR-----KPNLDVTQL-TTVCSDF 279
Qy 1103 SNAVS--LLSHLWLANATAALRINTDLLPTSQYLVKFEIVHFPQKTFERIFAFIGIPL- 1159
Db 280 SNSVSTGLMRPPLKAG-----KYMVLVRYEDLARNPMKTEIIVGFLGIPLD 325
Qy 1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTNVWKNLPRDEIKLIENI 1207
Db 326 SHVARWIONNTRGDPITLGHKHYGTVRN-----SAATAEKWRFLSYDIVAFAPNA 375
Qy 1208 CWTLMORLGY 1217
Db 376 CQOVLAQLGY 385

RESULT 11

US-10-408-765A-395
; Sequence 395, Application US/10408765A

```
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408.765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 411
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-408-765A-395

Query Match      2.3%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY      813 WILVIALWFIEL-----LDVWSTC-----SQPICKAKWTRTEAGSKKSLSS 855
Db      5 WKAVLLALASIAIOYTAIRTTAKSFHTCPLAELAGLAERLC-----ESPTFAYNLSR 59
QY      856 EGHMDLPDVVITSLPGSGAELKQLFPNSSFDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db      60 KTH-----ILILATTRSGSFVQLFNQHLDFVYLFEPL----- 93
QY      916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHPEPNRGKLAQYFAMNKDKRKFRRR 975
Db      94 -----YHVQNTLIIPRTQG-----KSPADRRV 115
QY      976 SLPQRQSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db      116 MLAGSRDLRLSLYDCDLYFLENYIKPPPVNHTDRIFFRGASRVLCSPVCDPPGADLV 175
QY      1014 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 1042
Db      176 LEEGDCVRKCGLLNLTVAEAACRERSHVAIKTVRVPEVNDLRALVEDPRNLKVIQLVRD 235
QY      1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYELPKELSKSK 1102
Db      236 PRGILASRSETFRDYY-----RLWRLWYGTGR-----KPYNLDVTQL-TTVCEDF 279
QY      1103 SNAYS--LLSHLWLANTAALRINTDLLPTSQYOLVKFEDIVHPQKTTTERIFAFGLIPL- 1159
Db      280 SNSVSTGLMRPPLWLG-----KMYLVRYEDLARNPMKKTTEEYIGFLGIPLD 325
QY      1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENI 1207
Db      326 SHVARWIONTRGDPFLGKHKGTVRN-----SAATAEKWRFRLSYDIVAFAQNA 375
QY      1208 CWTLMDRIGY 1217
Db      376 CQVLAQLGY 385

RESULT 13
US-09-817-514A-8
; Sequence 8, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 2504

; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408.765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 395
; LENGTH: 411
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-408-765A-395

Query Match      2.3%; Score 147; DB 16; Length 411;
Best Local Similarity 18.2%; Pred. No. 0.0008;
Matches 89; Conservative 58; Mismatches 149; Indels 194; Gaps 16;

QY      813 WILVIALWFIEL-----LDVWSTC-----SQPICKAKWTRTEAGSKKSLSS 855
Db      5 WKAVLLALASIAIOYTAIRTTAKSFHTCPLAELAGLAERLC-----ESPTFAYNLSR 59
QY      856 EGHMDLPDVVITSLPGSGAELKQLFPNSSFDFLYIRVPTAYIDIPETELEIDSFVDACE 915
Db      60 KTH-----ILILATTRSGSFVQLFNQHLDFVYLFEPL----- 93
QY      916 WKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHPEPNRGKLAQYFAMNKDKRKFRRR 975
Db      94 -----YHVQNTLIIPRTQG-----KSPADRRV 115
QY      976 SLPQRQSMKGAFDRDAEYI-----RALRRHLVYPSARPV-----LS 1013
Db      116 MLAGSRDLRLSLYDCDLYFLENYIKPPPVNHTDRIFFRGASRVLCSPVCDPPGADLV 175
QY      1014 LSSGSWTLKLHFFQEVLGA-----SMRALYIVRD 1042
Db      176 LEEGDCVRKCGLLNLTVAEAACRERSHVAIKTVRVPEVNDLRALVEDPRNLKVIQLVRD 235
QY      1043 PRAWIYMLYNSKPSLYSLKNVPEHLAKLFKIEGKGKCNLNSGYAFEYELPKELSKSK 1102
Db      236 PRGILASRSETFRDYY-----RLWRLWYGTGR-----KPYNLDVTQL-TTVCEDF 279
QY      1103 SNAYS--LLSHLWLANTAALRINTDLLPTSQYOLVKFEDIVHPQKTTTERIFAFGLIPL- 1159
Db      280 SNSVSTGLMRPPLWLG-----KMYLVRYEDLARNPMKKTTEEYIGFLGIPLD 325
QY      1160 -----SPASLNQILFATSTNLFYLPYEGEISPTNTNVKQNLPRDEIKLIENI 1207
Db      326 SHVARWIONTRGDPFLGKHKGTVRN-----SAATAEKWRFRLSYDIVAFAQNA 375
QY      1208 CWTLMDRIGY 1217
Db      376 CQVLAQLGY 385

RESULT 12
US-10-723-860-1544
; Sequence 1544, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
```


PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/007,255
 FILING DATE: 06-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/608,423
 FILING DATE: 28-FEB-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/705,484
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.93804
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2504 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-10-262-794A-12

Query Match 2.2%; Score 145; DB 15; Length 2504;
 Best Local Similarity 18.1%; Pred. No. 0.024;
 Matches 231; Conservative 180; Mismatches 443; Indels 422; Gaps 60;

QY 29 FSTFEESVSNSEWA-----VFTDDIDQFKTKQVDFRNQKLKSMHLPSLYFD 78
 DB 1003 FTDWERYNKYSTWAGSELVYVYENYDPTQRIQGTQKMD-ALLQSIQSLNADTVED 1061
 QY 79 AGEIQAMQKRSASHLHLFAIRSAVTV-----MLSNPTVYLPPEKHA-----D 122
 DB 1062 AFKTY-LTSFEQVANLKVISAHDNVNVDQGLTVYIGIDQAAPGYIYRVSVDHSCENGK 1120
 QY 123 FAA-----KWNIEYGNLPLALYCLLCPEDKVAPEFVLEYMDRMVGYKDWLNAAPGDEV 178
 DB 1121 FAANANGEMNKI-----TCVNPWKNIIRPVV--YNSRL--YLLMLEQSKSD- 1165
 QY 179 PIGSLTGFATAPDFLNLNDRHQKYLEKIW-----VITEMEYYSKVRSGWKOLH 232
 DB 1166 -----DGKTTIYQ--YNLKLHTR-----YDGSWNTPTFDVTEKVKNYTSSTDAAESL-- 1212
 QY 233 NHQATNMIALTGLVGTGVDKSKANIWKQAVVDVMEKTMFLNHIVDGSLDEGVAQSV 292
 DB 1213 -----GLYCTGYQG-----EDTLVMPFYSMQSS-----YSSY 1239
 QY 293 T---AKSVTQVFLAQRHFNINLLD--NNMLKMHFWFYATLLPGFORTVGIADSNYNWF 347
 DB 1240 TDNNAPVTGLYIFADMSDDNTNAQATYWNNSY-----PQPTVMADPDS----- 1286
 QY 348 YGPESQLVFLDKFLKNGAGNLAAQQRKHKRPDGMVPSTAQKRWSTLHTEYIYDPQLT 407
 DB 1287 -----NKKVITRRVNRVYAEVDE-----IPSSV-----TSNSYNSWGDHSLT 1323
 QY 408 POPPADYGTAKIHFTFPWGVVYTGAGLPN-----TQNTFVS-FKSGKLGRAV 455
 DB 1324 -----MLYGSVENITFESAEDLRLSTNMALSIITHNGYAGTRRI 1363
 QY 456 YDIVHFQPSWIDGRGFPNGHPDQNSFTFAPNGQVFVSEALYGPKLSHLNNLVFAP 515
 DB 1364 QCNLMKQVASLGDFFIYDSFD--DANRNLVP--LF-----KFGDENSDDSIICYNE 1414
 QY 516 SPSSQCNKPWBGQLGECAQMLKWTGE-----EVGDAAGEIITASQHGEMVFSGEAVSAYS 571
 DB 1415 NPSSE-DKKWYFSSKDDNKADYNGGTQCIDAGTSNKDFYVNLQIEIVISVTGGYSSYK 1473
 QY 572 SAMRLKSVYRALLLNSQTLVVDHIERQEDSPINSVSAFHNLDIDFKYIPKFMRYN 631
 DB 1474 IS-----NPNI-----NTGIDSAKV--KVTYKAG 1496
 QY 632 GAMDVMDAHYKMFDFDHGNSPWSAQEABQAAEFKKRWTFQVNVTFQWEST--ITRIA 689

DB 1497 G-----DDQIFTADNSTVYPQ-----QAPSPFEEMIYQFNLLTIDCKNLNFDNOA 1542
 QY 690 YVYFPGYINVSSCRFDIDSSNPGLOISLNVNNTHEVVSIVTDYHNL-----KTRFNVL 741
 DB 1543 HIEIDFTATAQDGRFLGAEFTFIIPVTYKVLGTENGVYSENNGVOYMOIGAYRRLNLT 1602
 QY 742 GFEGFASVADQG-----QITRFGLGTQAIKVPVRHRIIFPFGFKFNIAVGLI 789
 DB 1603 FAQQLVSRANRIGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGINKSKFALEYVDI 1662
 QY 790 LCISLVILTFQWRFLYSLFRKLMRWILLIIVIAMWFIELL-----DVMSTCSQPICAKW---- 841
 DB 1663 FKENDSFVIYQGLSETSTQTVVK-----VFLSYFIEATGKNKHLW-----VRKYQKET 1711
 QY 842 -----TRTEAE-----GSKSLSSSEGHMDLDPDVITSLPGSGA----- 875
 DB 1712 TDKILFDRTDEKDPHGWFSLDDHKFTFGLSSAQAALKNDSPEMDP-----SGANAL 1761
 QY 876 -----EILKOLFNFSSD--FLYIRVPTAVIDIPETETELEIDSVPDACEWVK 918
 DB 1762 YFWELFYPTMMMAHRLLEQNFDAANHFVRVWSPSGYI-----VDGKIAYHNVV 1813
 QY 919 SDIRSGHFLRLRGWLQSLVQDTKLJLQNIHLHPN-----RGKLAQYFA-----MN 964
 DB 1814 -----RPLEEDTSWNAQQLDSDTPDAVADDPMKYKVATFMATLDLLMAR 1858
 QY 965 KKKKPKFKRESLPEORSQMKGAFDRDAEYIYALR-----RHLVYYPSPARPVLSLSSGS 1018
 DB 1859 GDAAYRQLERDLAEAKMW-----YTQALNLLGDPEQVMLSTTWTANPTLGNAASK 1908
 QY 1019 WTLKLHFFOEVLGASNRALYIVRDPRAWIYSLMYSLKPSLSYKKNVPEHLAKLFKIEGK 1078
 DB 1909 TTQQVR--QQVL--TQRLNSRVKTP-----LLGTANSLTAL--FLPQENSKL-----K 1951
 QY 1079 G-----KCNLNGSYAFEPYERKELSKSNVSLSHLWLANAALINTDLP 1130
 DB 1952 GYWRTLAQRMFNLRHNLSDGQPLSLYAKPADPKALLSAVASQOGA----- 2001
 QY 1131 SYQLVKFEDIVH-FPQ 1145
 DB 2002 --DLPKAPLTIHRFPQ 2015

RESULT 15
 US-10-754-115-59
 ; Sequence 59, Application US/10754115
 ; Publication No. US20040208907A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hey, Timothy
 ; APPLICANT: Schleper, Amanda
 ; APPLICANT: Bevan, Scott
 ; APPLICANT: Bintrim, Scott
 ; APPLICANT: Mitchell, Jon
 ; APPLICANT: Li, Ze Sheng
 ; APPLICANT: Ni, Weiting
 ; APPLICANT: Zhu, Baolong
 ; APPLICANT: Merlo, Don
 ; APPLICANT: Apel-Birkhold, Patricia
 ; APPLICANT: Meade, Thomas
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
 ; FILE REFERENCE: DAS-104XCI
 ; CURRENT APPLICATION NUMBER: US/10/754,115
 ; CURRENT FILING DATE: 2004-01-07
 ; PRIOR APPLICATION NUMBER: US 60/441,723
 ; PRIOR FILING DATE: 2003-01-21
 ; NUMBER OF SEQ ID NOS: 64
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 59
 ; LENGTH: 2504
 ; TYPE: PRT
 ; ORGANISM: Photorhabdus luminescens
 ; US-10-754-115-59

Qy	919	SDIRSCHPELLRGWLOSLQVODFKLHLONTHLHEPN-----RCKLAQYPA-----MN	964
Db	1814	-----RPLEEDTSWNAQQDSDSTDPDAVAQDDPMHYKVATFWATLDDLMAW	1858
Qy	965	KDKCRKFKERESLPQRSQMKGAFFORDAEYIRALR-----RHLVYYPSPARFVLSLSSGS	1018
Db	1859	GDAAYRQLERDITLAEAKOW-----YTQALNLLGDPEQVWMLSTTWANPTLGNAAASK	1908
Qy	1019	WTLKULHFFOEVLGASWRALIYIVRDPAMIYSLMYNSKPSLYLSKVPHEHLAKLFLIEGGK	1078
Db	1909	TTQOVR--QOVL-TOLRLNSRVKTP-----LLGTANSLTAL-FLPQENSKL-----K	1951
Qy	1079	G-----KCNLNSGYAFEYBPRLKELSKSNVAVLSLHSLHWLANTAANALINTDLLPT	1130
Db	1952	GWVRTLAQRMFNRHNLSDGQFLSPLYAKPADPKALLSAVASAQGA-----	2001
Qy	1131	SYQLVKFEDIVH-FPQ	1145
Db	2002	--DLPKAPLTIHRFPQ	2015

Search completed: June 23, 2005, 09:28:02
Job time : 159.482 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:38 ; Search time 27.7171 Seconds.

(without alignments)
4242.027 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWIMALMFTGHL.....LIENICWTLMRLGYPKFM 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_79:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	251.5	3.9	672	2 T36083	hypothetical prote
2	192	2.9	776	2 D98354	oligo alginate lya
3	192	2.9	776	2 A2928	oligo alginate lya
4	133	2.0	484	2 J50261	N-acetylglucosamin
5	127	1.9	1456	1 A36563	mannose receptor p
6	125	1.9	4344	1 A53489	dynein heavy chain
7	123	1.9	1044	2 B72338	galactosyltransfer
8	122	1.9	1747	2 T43162	vitellogenin - gyp
9	120.5	1.8	944	2 D82926	hypothetical prote
10	113.5	1.7	484	2 JC7350	N-acetylglucosamin
11	113.5	1.7	3744	2 S46715	hypothetical prote
12	113	1.7	2216	2 S78398	hypothetical prote
13	112.5	1.7	4199	2 S76412	hypothetical prote
14	111	1.7	921	2 A51698	isooleucyl-tRNA syn
15	110.5	1.7	1248	2 C98974	autolysin [importe
16	110.5	1.7	1455	1 A48925	mannose receptor p
17	110	1.7	542	2 S54260	bromodomain protei
18	110	1.7	2095	2 S29529	genome polyprotein
19	110	1.7	3655	2 T38084	TRAP-like protein
20	109.5	1.7	1441	2 T00335	hypothetical prote
21	109	1.7	2401	2 T28676	rhostry protein -
22	108.5	1.7	866	2 T06454	probable lipoxigen
23	108.5	1.7	936	2 T1862	probable cytochrom
24	108.5	1.7	1215	2 T43916	chitinase A [limpor
25	108	1.7	1102	2 S44772	C29E4.4 protein -
26	108	1.7	1589	2 T22668	hypothetical prote
27	108	1.7	2076	2 S15999	fatty-acyl-CoA syn
28	107	1.6	771	2 H84845	hypothetical prote
29	107	1.6	1131	2 F82875	hypothetical prote

30 107 1.6 1816 2 A84845 probable ABC trans
31 107 1.6 1969 2 T38495 hypothetical prote
32 106.5 1.6 532 2 H96840 hypothetical prote
33 106.5 1.6 2619 2 T24588 hypothetical prote
34 106 1.6 486 2 JC7351 N-acetylglucosamin
35 106 1.6 921 2 AC1327 isooleucyl-tRNA syn
36 105.5 1.6 518 2 B83472 hypothetical prote
37 105.5 1.6 906 1 IJXIC2 N-cadherin 2 precu
38 105 1.6 386 2 S58755 ribosomal protein
39 105 1.6 877 1 IJBOCN N-cadherin precurs
40 105 1.6 910 2 T22050 hypothetical prote
41 105 1.6 924 2 T22044 hypothetical prote
42 104.5 1.6 586 2 D90250 glycogen synthase
43 104.5 1.6 664 2 D82056 ABC transporter, A
44 104.5 1.6 686 1 ALDYAT amylase A (EC 3.2.
45 104.5 1.6 929 2 T17392 .vrlL protein - Dic

ALIGNMENTS

RESULT 1

T36083

hypothetical protein SCE134.01c - Streptomyces coelicolor (fragment)

C:Species: Streptomyces coelicolor

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 12-Jul-2004

C:Accession: T36083

R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: Z21596

A:Accession: T36083

A:Status: preliminary; translated from GE/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-672 <SAU>

A:Cross-references: UNIPROT:Q8CJW4; EMBL:AL049661; PIDN:CAB41199.1; GSPDB:GN00070; SCOEDB:

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SCE134.01c

C:Superfamily: oligosaccharide lyase

Query Match 3.9%; Score 251.5; DB 2; Length 672;

Best Local Similarity 22.6%; Pred. No. 9.5e-10;

Matches 125; Conservative 86; Mismatches 249; Indels 93; Gaps 27;

Qy	80	GEIQMRQKSRAGSHLHFRAIRSAVTVMLSNPTVYL---PPPKHADFAAKWNEYGNLNP	136
Db	6	GRIDSLRGDLNGTHAAQWRRLHEQC-----DWYRRQNPFTEPEASITY---FGPAAA	55
Qy	137	PLALYCLLCPEDKVAFEFVLEYMDRMVGKDWLVENAPGDEVPIGHSLTGTFATPFDLYN	196
Db	56	NLALAYRLTQQ-RGYLLEARNWISTCVAYPHWGRAHMPDHDLDAGWLLHGLSLAYSWLGE	114
Qy	197	LLNHRROKYLEKIWIWTEMEYSKV---RSWCKQLLHHQATNMIALLTGALVTGYDK	253
Db	115	DLEPERREILRAKLELOGERLHSAFETTCGRWSSAYWQNH---NWIC-WTGATAGYAL	170
Qy	254	GSRANIWKQAVDVDMKTMFLNHHIVDGLDEGVAGSYTAKSVTVQVFLAQRH-----	307
Db	171	GRSE--WTKAARANLETULTMLPE--DGSSEGVVYWRVGVFWLAIHTDLVQREQADLM	226
Qy	308	----FNINLNDNLWKHFWFYATLPGFQRTVGIADSNYNWFYGPESQLVFLDKFLK	363
Db	227	STGGFLNT--TRW-RLH-----QSAPGFEENIDHGDC-HDRSGHSHVALYRLASAYQ	276
Qy	364	NGAGNWLAQOI-RKH-----RPNQGMVPSAQRMSTLHTEYIWDPQITP-OPP	411
Db	277	DGTAQMLGNLVAERHFWREAYESGVRPGVMP-----EAFLELLWYDPRVTPAAPD	326
Qy	412	ADVGTAKIHTFNNWVVTYAGLPTNTTFVSPKSGKLGGRVYDVIVHFQFYSWIDQWR	471
Db	327	REPQTA---YFPDUGQITARTGWDNSAA--TCVSPKAAFGGHRADWDGH--RLKAAGWD	379
Qy	472	SFNPGEHPDQNSFTFAPNGQVFVSEALYG--PKLSHLNNVLNVFAPSPSSQCNKFWEQOL	529

Qy 442 FVGFKSGKGLGRAVYDIYVHFQPYSWIDGRSFRNPGHEHPDQNSFTFAPNGQVVFSEALY 501
 Db 519 FV-FKS-----SPYSL-----SHSGDQNAFVLAHGEDLAIQSGY- 554
 Qy 502 PKLSHLNNVLVFPSPSQCKPWEQ-----LGECAQWLKWTGEEVGDAAAGEIITA 553
 Db 555 -----YVAF-----NSQHLNRRQTRSKNAVLIGKGQYAEKQKALARRAAGRIVS 602
 Qy 554 SQHCEMVFVSGEAVSAYSSAMRL-KSVYRALLLNSQTLVVDHIERQE----- 601
 Db 603 EQQGHVRIYVGDATAAIVANPLVQKVLRETHFVNDYFYVIDVEVCSBQELQWLCHTL 662
 Qy 602 DSPINSVAFPHNLDIDFKYIPKFNRYNGAMVDVAHYKMFVFDHGHGNSPMASTQ 659
 Db 663 GAPQGRSSF-----RYNGRKA-----GFYGVQFVSSGGTTPQISAVE 699

RESULT 4
 JE0261
 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
 C:Species: Homo sapiens (man)
 C>Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
 C:Accession: JE0261
 R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.
 J. Biochem. 124, 670-678, 1998
 A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
 A:Reference number: JE0261; MUID:98391845; PMID:9722682
 A:Accession: JE0261
 A:Molecule type: mRNA
 A:Residues: 1-484 <UCH>
 A:Cross-references: DDBJ:AB014679
 C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine
 C:Superfamily: chondroitin 6-sulfotransferase
 C:Keywords: sulfotransferase

Query Match 2.0%; Score 133; DB 2; Length 484;
 Best Local Similarity 26.2%; Pred. No. 0.18;
 Matches 53; Conservative 36; Mismatches 81; Indels 32; Gaps 9;

Qy 1034 MRALYIVRDPRAWYSLMYNSKPSLY--SLKNVPE-----HLAKLFKIEGK-GKCNLS 1085
 Db 279 LKVIHLVDRPA-VASSRIRSRHGLRESLOVVRSDPRAPRAHMPFLEAAGHLGAKKEGV 337
 Qy 1086 GYAFYEPRLKELSKSNVSLSHLWLANAALINTDLPSTVOLVKFEDIVHPQ 1145
 Db 338 GGPADYHAL-----GAMEVICNSMAKTLQALQ-PPDWLQGHVLYVRYEDLVGEPV 387
 Qy 1146 KTTERTAFGLIPLSA-----SLNQILPATSNLFPYLYEIGEISPTN-----TNVWKQNL 1197
 Db 388 KTLRRVYDFVGLLVSPMEQFALNWTSGSGSKPFV-----VSARNATQANAWRTALT 442

Qy 1198 RDEIKLIENICWTMLDRLLGYPK 1219
 Db 443 FQKIQVEEFCYQPAVLGYER 464

RESULT 5
 A36563
 mannose receptor precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C:Accession: A36563; A60926; A44255; B44255; C44255; D44255; F44255; G44255; H44255; I44255; J44255; K44255; L44255; M44255; N44255; O44255; P44255; Q44255; R44255; S44255; T44255; U44255; V44255; W44255; X44255; Y44255; Z44255; AA44255; AB44255; AC44255; AD44255; AE44255; AF44255; AG44255; AH44255; AI44255; AJ44255; AK44255; AL44255; AM44255; AN44255; AO44255; AP44255; AQ44255; AR44255; AS44255; AT44255; AU44255; AV44255; AW44255; AX44255; AY44255; AZ44255; BA44255; BB44255; BC44255; BD44255; BE44255; BF44255; BG44255; BH44255; BI44255; BJ44255; BK44255; BL44255; BM44255; BN44255; BO44255; BP44255; BQ44255; BR44255; BS44255; BT44255; BU44255; BV44255; BW44255; BX44255; BY44255; BZ44255; CA44255; CB44255; CC44255; CD44255; CE44255; CF44255; CG44255; CH44255; CI44255; CJ44255; CK44255; CL44255; CM44255; CN44255; CO44255; CP44255; CQ44255; CR44255; CS44255; CT44255; CU44255; CV44255; CW44255; CX44255; CY44255; CZ44255; DA44255; DB44255; DC44255; DD44255; DE44255; DF44255; DG44255; DH44255; DI44255; DJ44255; DK44255; DL44255; DM44255; DN44255; DO44255; DP44255; DQ44255; DR44255; DS44255; DT44255; DU44255; DV44255; DW44255; DX44255; DY44255; DZ44255; EA44255; EB44255; EC44255; ED44255; EE44255; EF44255; EG44255; EH44255; EI44255; EJ44255; EK44255; EL44255; EM44255; EN44255; EO44255; EP44255; EQ44255; ER44255; ES44255; ET44255; EU44255; EV44255; EW44255; EX44255; EY44255; EZ44255; FA44255; FB44255; FC44255; FD44255; FE44255; FG44255; FH44255; FI44255; FJ44255; FK44255; FL44255; FM44255; FN44255; FO44255; FP44255; FQ44255; FR44255; FS44255; FT44255; FU44255; FV44255; FW44255; FX44255; FY44255; FZ44255; GA44255; GB44255; GC44255; GD44255; GE44255; GF44255; GG44255; GH44255; GI44255; GJ44255; GK44255; GL44255; GM44255; GN44255; GO44255; GP44255; GQ44255; GR44255; GS44255; GT44255; GU44255; GV44255; GW44255; GX44255; GY44255; GZ44255; HA44255; HB44255; HC44255; HD44255; HE44255; HF44255; HG44255; HH44255; HI44255; HJ44255; HK44255; HL44255; HM44255; HN44255; HO44255; HP44255; HQ44255; HR44255; HS44255; HT44255; HU44255; HV44255; HW44255; HX44255; HY44255; HZ44255; IA44255; IB44255; IC44255; ID44255; IE44255; IF44255; IG44255; IH44255; II44255; IJ44255; IK44255; IL44255; IM44255; IN44255; IO44255; IP44255; IQ44255; IR44255; IS44255; IT44255; IU44255; IV44255; IW44255; IX44255; IY44255; IZ44255; JA44255; JB44255; JC44255; JD44255; JE44255; JF44255; JG44255; JH44255; JI44255; JJ44255; JK44255; JL44255; JM44255; JN44255; JO44255; JP44255; JQ44255; JR44255; JS44255; JT44255; JU44255; JV44255; JW44255; JX44255; JY44255; JZ44255; KA44255; KB44255; KC44255; KD44255; KE44255; KF44255; KG44255; KH44255; KI44255; KJ44255; KK44255; KL44255; KM44255; KN44255; KO44255; KP44255; KQ44255; KR44255; KS44255; KT44255; KU44255; KV44255; KW44255; KX44255; KY44255; KZ44255; LA44255; LB44255; LC44255; LD44255; LE44255; LF44255; LG44255; LH44255; LI44255; LJ44255; LK44255; LL44255; LM44255; LN44255; LO44255; LP44255; LQ44255; LR44255; LS44255; LT44255; LU44255; LV44255; LW44255; LX44255; LY44255; LZ44255; MA44255; MB44255; MC44255; MD44255; ME44255; MF44255; MG44255; MH44255; MI44255; MJ44255; MK44255; ML44255; MM44255; MN44255; MO44255; MP44255; MQ44255; MR44255; MS44255; MT44255; MU44255; MV44255; MW44255; MX44255; MY44255; MZ44255; NA44255; NB44255; NC44255; ND44255; NE44255; NF44255; NG44255; NH44255; NI44255; NJ44255; NK44255; NL44255; NM44255; NO44255; NP44255; NQ44255; NR44255; NS44255; NT44255; NU44255; NV44255; NW44255; NX44255; NY44255; NZ44255; OA44255; OB44255; OC44255; OD44255; OE44255; OF44255; OG44255; OH44255; OI44255; OJ44255; OK44255; OL44255; OM44255; ON44255; OO44255; OP44255; OQ44255; OR44255; OS44255; OT44255; OU44255; OV44255; OW44255; OX44255; OY44255; OZ44255; PA44255; PB44255; PC44255; PD44255; PE44255; PF44255; PG44255; PH44255; PI44255; PJ44255; PK44255; PL44255; PM44255; PN44255; PO44255; PP44255; PQ44255; PR44255; PS44255; PT44255; PU44255; PV44255; PW44255; PX44255; PY44255; PZ44255; QA44255; QB44255; QC44255; QD44255; QE44255; QF44255; QG44255; QH44255; QI44255; QJ44255; QK44255; QL44255; QM44255; QN44255; QO44255; QP44255; QQ44255; QR44255; QS44255; QT44255; QU44255; QV44255; QW44255; QX44255; QY44255; QZ44255; RA44255; RB44255; RC44255; RD44255; RE44255; RF44255; RG44255; RH44255; RI44255; RJ44255; RK44255; RL44255; RM44255; RN44255; RO44255; RP44255; RQ44255; RR44255; RS44255; RT44255; RU44255; RV44255; RW44255; RX44255; RY44255; RZ44255; SA44255; SB44255; SC44255; SD44255; SE44255; SF44255; SG44255; SH44255; SI44255; SJ44255; SK44255; SL44255; SM44255; SN44255; SO44255; SP44255; SQ44255; SR44255; SS44255; ST44255; SU44255; SV44255; SW44255; SX44255; SY44255; SZ44255; TA44255; TB44255; TC44255; TD44255; TE44255; TF44255; TG44255; TH44255; TI44255; TJ44255; TK44255; TL44255; TM44255; TN44255; TO44255; TP44255; TQ44255; TR44255; TS44255; TT44255; TU44255; TV44255; TW44255; TX44255; TY44255; TZ44255; UA44255; UB44255; UC44255; UD44255; UE44255; UF44255; UG44255; UH44255; UI44255; UJ44255; UK44255; UL44255; UM44255; UN44255; UO44255; UP44255; UQ44255; UR44255; US44255; UT44255; UU44255; UV44255; UW44255; UX44255; UY44255; UZ44255; VA44255; VB44255; VC44255; VD44255; VE44255; VF44255; VG44255; VH44255; VI44255; VJ44255; VK44255; VL44255; VM44255; VN44255; VO44255; VP44255; VQ44255; VR44255; VS44255; VT44255; VU44255; VV44255; VW44255; VX44255; VY44255; VZ44255; WA44255; WB44255; WC44255; WD44255; WE44255; WF44255; WG44255; WH44255; WI44255; WJ44255; WK44255; WL44255; WM44255; WN44255; WO44255; WP44255; WQ44255; WR44255; WS44255; WT44255; WU44255; WV44255; WW44255; WX44255; WY44255; WZ44255; XA44255; XB44255; XC44255; XD44255; XE44255; XF44255; XG44255; XH44255; XI44255; XJ44255; XK44255; XL44255; XM44255; XN44255; XO44255; XP44255; XQ44255; XR44255; XS44255; XT44255; XU44255; XV44255; XW44255; XX44255; XY44255; XZ44255; YA44255; YB44255; YC44255; YD44255; YE44255; YF44255; YG44255; YH44255; YI44255; YJ44255; YK44255; YL44255; YM44255; YN44255; YO44255; YP44255; YQ44255; YR44255; YS44255; YT44255; YU44255; YV44255; YW44255; YX44255; YY44255; YZ44255; ZA44255; ZB44255; ZC44255; ZD44255; ZE44255; ZF44255; ZG44255; ZH44255; ZI44255; ZJ44255; ZK44255; ZL44255; ZM44255; ZN44255; ZO44255; ZP44255; ZQ44255; ZR44255; ZS44255; ZT44255; ZU44255; ZV44255; ZW44255; ZX44255; ZY44255; ZZ44255; AA44255; AB44255; AC44255; AD44255; AE44255; AF44255; AG44255; AH44255; AI44255; AJ44255; AK44255; AL44255; AM44255; AN44255; AO44255; AP44255; AQ44255; AR44255; AS44255; AT44255; AU44255; AV44255; AW44255; AX44255; AY44255; AZ44255; BA44255; BB44255; BC44255; BD44255; BE44255; BF44255; BG44255; BH44255; BI44255; BJ44255; BK44255; BL44255; BM44255; BN44255; BO44255; BP44255; BQ44255; BR44255; BS44255; BT44255; BU44255; BV44255; BW44255; BX44255; BY44255; BZ44255; CA44255; CB44255; CC44255; CD44255; CE44255; CF44255; CG44255; CH44255; CI44255; CJ44255; CK44255; CL44255; CM44255; CN44255; CO44255; CP44255; CQ44255; CR44255; CS44255; CT44255; CU44255; CV44255; CW44255; CX44255; CY44255; CZ44255; DA44255; DB44255; DC44255; DD44255; DE44255; DF44255; DG44255; DH44255; DI44255; DJ44255; DK44255; DL44255; DM44255; DN44255; DO44255; DP44255; DQ44255; DR44255; DS44255; DT44255; DU44255; DV44255; DW44255; DX44255; DY44255; DZ44255; EA44255; EB44255; EC44255; ED44255; EE44255; EF44255; EG44255; EH44255; EI44255; EJ44255; EK44255; EL44255; EM44255; EN44255; EO44255; EP44255; EQ44255; ER44255; ES44255; ET44255; EU44255; EV44255; EW44255; EX44255; EY44255; EZ44255; FA44255; FB44255; FC44255; FD44255; FE44255; FG44255; FH44255; FI44255; FJ44255; FK44255; FL44255; FM44255; FN44255; FO44255; FP44255; FQ44255; FR44255; FS44255; FT44255; FU44255; FV44255; FW44255; FX44255; FY44255; FZ44255; GA44255; GB44255; GC44255; GD44255; GE44255; GF44255; GG44255; GH44255; GI44255; GJ44255; GK44255; GL44255; GM44255; GN44255; GO44255; GP44255; GQ44255; GR44255; GS44255; GT44255; GU44255; GV44255; GW44255; GX44255; GY44255; GZ44255; HA44255; HB44255; HC44255; HD44255; HE44255; HF44255; HG44255; HH44255; HI44255; HJ44255; HK44255; HL44255; HM44255; HN44255; HO44255; HP44255; HQ44255; HR44255; HS44255; HT44255; HU44255; HV44255; HW44255; HX44255; HY44255; HZ44255; IA44255; IB44255; IC44255; ID44255; IE44255; IF44255; IG44255; IH44255; II44255; IJ44255; IK44255; IL44255; IM44255; IN44255; IO44255; IP44255; IQ44255; IR44255; IS44255; IT44255; IU44255; IV44255; IW44255; IX44255; IY44255; IZ44255; JA44255; JB44255; JC44255; JD44255; JE44255; JF44255; JG44255; JH44255; JI44255; JJ44255; JK44255; JL44255; JM44255; JN44255; JO44255; JP44255; JQ44255; JR44255; JS44255; JT44255; JU44255; JV44255; JW44255; JX44255; JY44255; JZ44255; KA44255; KB44255; KC44255; KD44255; KE44255; KF44255; KG44255; KH44255; KI44255; KJ44255; KK44255; KL44255; KM44255; KN44255; KO44255; KP44255; KQ44255; KR44255; KS44255; KT44255; KU44255; KV44255; KW44255; KX44255; KY44255; KZ44255; LA44255; LB44255; LC44255; LD44255; LE44255; LF44255; LG44255; LH44255; LI44255; LJ44255; LK44255; LL44255; LM44255; LN44255; LO44255; LP44255; LQ44255; LR44255; LS44255; LT44255; LU44255; LV44255; LW44255; LX44255; LY44255; LZ44255; MA44255; MB44255; MC44255; MD44255; ME44255; MF44255; MG44255; MH44255; MI44255; MJ44255; MK44255; ML44255; MM44255; MN44255; MO44255; MP44255; MQ44255; MR44255; MS44255; MT44255; MU44255; MV44255; MW44255; MX44255; MY44255; MZ44255; NA44255; NB44255; NC44255; ND44255; NE44255; NF44255; NG44255; NH44255; NI44255; NJ44255; NK44255; NL44255; NM44255; NO44255; NP44255; NQ44255; NR44255; NS44255; NT44255; NU44255; NV44255; NW44255; NX44255; NY44255; NZ44255; OA44255; OB44255; OC44255; OD44255; OE44255; OF44255; OG44255; OH44255; OI44255; OJ44255; OK44255; OL44255; OM44255; ON44255; OO44255; OP44255; OQ44255; OR44255; OS44255; OT44255; OU44255; OV44255; OW44255; OX44255; OY44255; OZ44255; PA44255; PB44255; PC44255; PD44255; PE44255; PF44255; PG44255; PH44255; PI44255; PJ44255; PK44255; PL44255; PM44255; PN44255; PO44255; PP44255; PQ44255; PR44255; PS44255; PT44255; PU44255; PV44255; PW44255; PX44255; PY44255; PZ44255; QA44255; QB44255; QC44255; QD44255; QE44255; QF44255; QG44255; QH44255; QI44255; QJ44255; QK44255; QL44255; QM44255; QN44255; QO44255; QP44255; QQ44255; QR44255; QS44255; QT44255; QU44255; QV44255; QW44255; QX44255; QY44255; QZ44255; RA44255; RB44255; RC44255; RD44255; RE44255; RF44255; RG44255; RH44255; RI44255; RJ44255; RK44255; RL44255; RM44255; RN44255; RO44255; RP44255; RQ44255; RR44255; RS44255; RT44255; RU44255; RV44255; RW44255; RX44255; RY44255; RZ44255; SA44255; SB44255; SC44255; SD44255; SE44255; SF44255; SG44255; SH44255; SI44255; SJ44255; SK44255; SL44255; SM44255; SN44255; SO44255; SP44255; SQ44255; SR44255; SS44255; ST44255; SU44255; SV44255; SW44255; SX44255; SY44255; SZ44255; TA44255; TB44255; TC44255; TD44255; TE44255; TF44255; TG44255; TH44255; TI44255; TJ44255; TK44255; TL44255; TM44255; TN44255; TO44255; TP44255; TQ44255; TR44255; TS44255; TT44255; TU44255; TV44255; TW44255; TX44255; TY44255; TZ44255; UA44255; UB44255; UC44255; UD44255; UE44255; UF44255; UG44255; UH44255; UI44255; UJ44255; UK44255; UL44255; UM44255; UN44255; UO44255; UP44255; UQ44255; UR44255; US44255; UT44255; UU44255; UV44255; UW44255; UX44255; UY44255; UZ44255; VA44255; VB44255; VC44255; VD44255; VE44255; VF44255; VG44255; VH44255; VI44255; VJ44255; VK44255; VL44255; VM44255; VN44255; VO44255; VP44255; VQ44255; VR44255; VS44255; VT44255; VU44255; VV44255; VW44255; VX44255; VY44255; VZ44255; WA44255; WB44255; WC44255; WD44255; WE44255; WF44255; WG44255; WH44255; WI44255; WJ44255; WK44255; WL44255; WM44255; WN44255; WO44255; WP44255; WQ44255; WR44255; WS44255; WT44255; WU44255; WV44255; WW44255; WX44255; WY44255; WZ44255; XA44255; XB44255; XC44255; XD44255; XE44255; XF44255; XG44255; XH44255; XI44255; XJ44255; XK44255; XL44255; XM44255; XN44255; XO44255; XP44255; XQ44255; XR44255; XS44255; XT44255; XU44255; XV44255; XW44255; XX44255; XY44255; XZ44255; YA44255; YB44255; YC44255; YD44255; YE44255; YF44255; YG44255; YH44255; YI44255; YJ44255; YK44255; YL44255; YM44255; YN44255; YO44255; YP44255; YQ44255; YR44255; YS44255; YT44255; YU44255; YV44255; YW44255; YX44255; YY44255; YZ44255; ZA44255; ZB44255; ZC44255; ZD44255; ZE44255; ZF44255; ZG44255; ZH44255; ZI44255; ZJ44255; ZK44255; ZL44255; ZM44255; ZN44255; ZO44255; ZP44255; ZQ44255; ZR44255; ZS44255; ZT44255; ZU44255; ZV44255; ZW44255; ZX44255; ZY44255; ZZ44255; AA44255; AB44255; AC44255; AD44255; AE44255; AF44255; AG44255; AH44255; AI44255; AJ44255; AK44255; AL44255; AM44255; AN44255; AO44255; AP44255; AQ44255; AR44255; AS44255; AT44255; AU44255; AV44255; AW44255; AX44255; AY44255; AZ44255; BA44255; BB44255; BC44255; BD44255; BE44255; BF44255; BG44255; BH44255; BI44255; BJ44255; BK44255; BL44255; BM44255; BN44255; BO44255; BP44255; BQ44255; BR44255; BS44255; BT44255; BU44255; BV44255; BW44255; BX44255; BY44255; BZ44255; CA44255; CB44255; CC44255; CD44255; CE44255; CF44255; CG44255; CH44255; CI44255; CJ44255; CK44255; CL44255; CM44255; CN44255; CO44255; CP44255; CQ44255; CR44255; CS44255; CT44255; CU44255; CV44255; CW44255; CX44255; CY44255; CZ44255; DA44255; DB44255; DC44255; DD44255; DE44255; DF44255; DG44255; DH44255; DI44255; DJ44255; DK44255; DL44255; DM44255; DN44255; DO44255; DP44255; DQ44255; DR44255; DS44255; DT44255; DU44255; DV44255; DW44255; DX44255; DY44255; DZ44255; EA44255; EB44255; EC44255; ED44255; EE44255; EF44255; EG44255; EH44255; EI44255; EJ44255; EK44255; EL44255; EM44255; EN44255; EO44255; EP44255; EQ44255; ER44255; ES44255; ET44255; EU44255; EV44255; EW44255; EX44255; EY44255; EZ44255; FA44255; FB44255; FC44255; FD44255; FE44255; FG44255; FH44255; FI44255; FJ44255; FK44255; FL44255; FM44255; FN44255; FO44255; FP44255; FQ44255; FR44255; FS44255; FT44255; FU44255; FV44255; FW44255; FX44255; FY44

Db 295 HI-----NHKKDPAWMLQIIGMLRKIDERYTLIAGDFQ-----BIRYANYPKH----- 338
Qy 701 SCRFIDSSNPGLOISLNVNN--TEHVSIVTDTYHNLKTRFNVLGRF-GFASVADQOQITRF 758
Db 339 ---FI--KDAGLEKXVLYGWEDVNAFLDKXVLLSTSHESGYNIAEMAKG----- 388
Qy 759 GLGTQAIKVRPHD-----RIIFPGPKFNIAVGLICISLIVLTFQWRFY----- 804
Db 389 -----IKPIIHNAGAKTQWPDLLVFN-----IDEVIRIVTSRDYNSEKYSFVE 434
Qy 805 --LSPRKLMWILLVIALFIELLDVWSTCSQPICAKWTTEAGSKSLSSSEGHMDL 862
Db 435 KNCLEKQITSLIV-----QLDNNRTKNKSIHKST-TDTENSFAKIWEYSKID- 486
Qy 863 PDVVTSLPSGG-----AEILKOLF-FNSDFLIRVPTAY-----IDIPE 902
Db 487 SFTIMDLPGKLSRSEFVSLDERFIILNKAILVGTGAFSIELALREADVTGIDIDP 546
Qy 903 TELEI-----DSFVDACEKVSIRSGHFRLLR-----GWLQSLVQDTKLH 943
Db 547 TSIELAIRISKDYNVENVEPKVD-----GPKLTSPKQPEFDIAFNMGVVEHPKDDDIK 602
Qy 944 LONILHEPNRGLAQIAPAMNKDKRKFRRRESLPEORSOMKA-----FORDAEYIRALR 999
Db 603 MLK-----QMGVAKFVVGVYPYSGSFYKTA--KETAQKLGAWEGYGFERDFTLEPLI 654
Qy 1000 RHLVYVPSARPVLSSSGSWTLKLFHQEVLGASMRALYIVRDPRAVIYMLYNSKPSLY 1059
Db 655 RRAGLIP-----LH--EEVIGVLAEPFYLRRINPEW----- 684
Qy 1060 SLKXVPEHLAKLFKIE--GGKGNLNS--GYAEYEPPL--RKELSKSKNAVALLS 1110
Db 685 PLK-IAENLQYFOGEKVGWGLICFAITKWPYADEFLKLDHKKI-KFESQISLIT 739
RESULT 8
T43162
vitellogenin - gypsy moth
C:Species: Lymantria dispar (gypsy moth)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T43162; T43268
R:Hiremath, S.T.; Lehtoma, K.
submitted to the EMBL Data Library, June 1996
A:Title: Structure of the gypsy moth vitellogenin gene.
A:Reference number: 222378; MUID:97468497; PMID:9327584
A:Description: Gypsy moth vitellogenin mRNA.
A:Reference number: 222321
A:Accession: T43162
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1747 <HIR>
A:Cross-references: UNIPROT:Q25269; EMBL:U60186; NID:g1399940; PID:g1399941; PIDN:AA033
R:Hiremath, S.; Lehtoma, K.
Arch. Insect Biochem. Physiol. 36, 151-164, 1997
A:Title: Structure of the gypsy moth vitellogenin gene.
A:Reference number: 222378; MUID:97468497; PMID:9327584
A:Accession: T43268
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-819, 'Q', 821-1747 <H12>
A:Cross-references: EMBL:U90756; NID:G2894826; PID:G2894827; PIDN:AA02818.1
C:Genetics:
A:Introns: 11/1; 725/2; 796/3; 1120/1; 1410/1; 1477/2
C:Superfamily: boll weevil vitellogenin
Query Match 1.9%; Score 122; DB 2; Length 1747;
Best Local Similarity 20.4%; Pred. No. 8.2;
Matches 268; Conservative 174; Mismatches 536; Indels 336; Gaps 70;
Qy 36 VSNYSEWAVFTDDIDQ-----FKTKVQDFRPNQ---KUKSKMLH-----PS 74
Db 459 IAKVDMWFIFRDVIAQGTQAPQFIETWIKTIQGFLEAEVSSIAKTLRLHKTMTK 518
Qy 75 LYFDAGEI-QAMRQKSRASHLHLFRAIRSAVTVMLSNPTY-----YLPPEKHADFAAKWN 128

Db 519 QFFELALCPVLEQK-----YLNSSALITAAAEFIRLGSIDYEHSYPPYPPYVHGFPPLKND 574
Qy 129 E-YIGNNLPEPLA-LYCLLCPEDKVAPEFVLEYMDRMVGYKDWLVNENAPGDEVPIGHSLS-- 184
Db 575 RFIIENILPRLSQLLKQAVEQEDYGKTLVIVKAIGTLGHREILKVFTPYLEGQIKVSTYL 634
Qy 185 -TGFAFAFDLYNLLDNHRRQKYLEKIWIITEENWEYSKVSRSWKQLLH-NHQATNMIAL 242
Db 635 RVQMILALKPLAEQKORYVRSALFSILMTABPYE---VRVAAAMILFLNEPTTDLRV 690
Qy 243 LTGALVTGVDKGSKANIKWQAVDVMEKTMFLN-----HIVDG-----SLDEGVA 288
Db 691 M--AQLTND--PSIHVRAVLKSSITETAATLEGPKYWHLAKAAQSKVELVTSDFGFA 744
Qy 289 YG-----SYTAKSVTOYVFLAQRHENINNDNNMLKMHF---W---FYVATLLPGFOR 335
Db 745 YSAWPFAPKPFDEEVTRY-----RLANYIGSDGSLIPNLFSLTWNRRFYGR---GAEN 795
Qy 336 TVGIADSNYMWFGPESQLVFLDKFILKNGAGNWLAAQIRK-HRPKQGPMPV-STAQRWS 393
Db 796 TVGFTSSNVQDIPDIYIKQLILQPR--KSTTTEHQPSADKIAKLLNIKQDPLQPIESSIYK 853
Qy 394 TLHTEYIWDYDPLTPQPPADYGTAKIHTFPNWGV--TYGAGLPNTQNTTFVSFKSKL-- 450
Db 854 IFDKEFPW-----PLDQPNIKQLVNDVFGVKELENGVQSHYTKVFYSKQSVIFP 904
Qy 451 ---GGRAYVD---IVHQPYSWIDGWSRFPNGHEHPDQ-----SFTFAPNQVQVSE 497
Db 905 IATGEPFVFPKPVVHLQ--TKLSKINYPPTSKDISSNIQTEIOFTFARN-----ME 957
Qy 498 ALXGPKLSHLNVLVAFSPSSQCNKPWE---GQLGECAQWLKWTGEEVGDAAGEIITA 553
Db 958 GSVGFIDTNTNQLANIGVVKYQINLPKINVIGSGKFKVL-----EPLHT 1005
Qy 554 SQHGEVVFSGEAVSAYSSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVSAPFH 613
Db 1006 DQDLILLHSHWMPYSAYQOTNTTTPSHDLTKLKLKLVKVAEMKELGQOQFNLL----- 1060
Qy 614 NLDDIDKVIYKFMNRVNGA-----MMDVMDAHYKMFDFHGHGSPMASI 658
Db 1061 ---FXHGFSSNTYRNAGDFVQDADLDFSDPFMSLNDLAETHYE---FTVSGK----- 1106
Qy 659 QEAQAAEFKKRWI-----QFVNVY-----FQMESITIRIAYFYGYINV 699
Db 1107 QSKTKAVTTAAVDVSHNLSTGKFGAHRFSDVTPNSAPRREIVKRS-----SGIKT 1161
Qy 700 SSCRFID--SSNPGLOISLNVNNTEHVSVIVTDYHNLKTRFN-YLFGGPGFASVADQOQIT 756
Db 1162 ATARVIDFSASFEGLO-KLEYAVTAAVAGSMVD---LKTQPAVPMG-----SOSDNGOI- 1211
Qy 757 RFLGTOAIKVPVRHRIIFPGFKFNIAVGLICISLIVLTFQWRFVLSFR---KLWR 812
Db 1212 -----NAVFKLOKQPMADPFHKAALNSAVKYLEAD---VTYGENSININFGHTERSQE 1262
Qy 813 WILITVIALWFIELLDVWSTCSQPICAKWTTEAGSKSLSS-----EGHMDLPDVI 867
Db 1263 YAEQKNSLMA-----NQCAQ-----ENAAQNKFGQCHNLVIESHAPDRFKASI 1307
Qy 868 T--SLPGSGABILKQLFFN--SSDFLYIRVPTAVIDIPETELED-SFVD-----ACBWK 917
Db 1308 TYKHIPAAHTALLDSYIQLGWSKGFY--NPSKRLPVQIELEANASYVDQTANVAVTWM 1365
Qy 918 VSDIRSGHF-----RLRGWLSQSLVQDTKLHLON---IHLHPNRGKLAQYFAMNKDK 967
Db 1366 NGQVRFNLPNTYITPALTAYQPIGIEDSWTHFPANSYSHQYEP-----FCTVDGTK 1418
Qy 968 KRKFRRRE---SLPE-----QRSQMK-----GAFDRDAEYIR 996
Db 1419 VKTFSRDYVNTVPEIWTVLHQAQTNWELVVLAKRPNKAEKAKKEIGKIGKLDLYISHKT 1478
Qy 997 ALRHL--VYFSPARPVLSSSGSWTLKLFHQEVLGASMRALYIVRDPRAVIYMLYNSK 1055

Db 1479 ATGKHLEVNIP-----YSAANNKANVKVETNAQLVADGLTTY--WDDVAETPLLOYSNH 1531
Qy 1056 PSLYSLNKVNPEHLAKLFIKGG-----GKCNLSNGVAFYEPPLRKLSKSN 1104
Db 1532 PDRVLLHLSDGLHLLPDGKRGIFTTSQYRNITRGICGQNSG-----DPL--DDYKTPLG 1584
Qy 1105 AVSLLSHLWNTAAALRINTDLLPTSQYOLVKFEDI-----VHFPOKTERIFAF 1154
Db 1585 IVDHSQHFGAFT-----LDLEKNSQIQQWKKAQETAYQPKLTHVTILRF 1631

RESULT 9
D82926
hypothetical protein UUI166 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82926
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: D82926
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-944 <GLA>
A:Cross-references: GB:AE002116; GB:AF222894; NID:g6899118; PIDN:AAF30573.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UUI166
A:Genetic code: SGC3

Query Match 1.8%; Score 120.5; DB 2; Length 944;
Best Local Similarity 17.5%; Pred. No. 4;
Matches 212; Conservative 148; Mismatches 388; Indels 461; Gaps 58;

Qy 44 VFTDDIDQFTQKQVQD-----FRPNQKLKKSMLPSLYFDAGEIQAMRQKSRASHLHLFR 98
Db 9 IYLDLEGYSQKNIEIDSLSIEPNQK-----PIHLCPFR 43
Qy 99 AIRSAVTMLNSPTYLYLPPPKHADFAAKNEIYGNLPPALYCLICPEBK-----VAFEFV 155
Db 44 ILKGFIKITNTGQSLVNDK--DFSTVMN-IKASNEPIIAYW-----DKSPTYNFKGS 94
Qy 156 LEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATAPDFLYNLDNHRKQKLEKIWIITE 215
Db 95 LNVISK-----DYARIDPSNTLKNPF-----KPYYPQAWV--- 125
Qy 216 EMVEYSKVRSGKQLLHNHOATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFL 275
Db 126 --QNYQANNVSTIVNDKEHTWIKTITVTTLVLYKPSVVEDEIYQNE----- 171
Qy 276 NHIVDGLS-DEGVAYGSYTAQSYTVQYVF-LAQRHFNINLNDNMLKMHFWFYATLLPGF 333
Db 172 ---QGNLVSHKAKFKQWTTMDLKEVLIIPDSEAQAQ-----WLMVNEVIPEY 217
Qy 334 QRTVGIADSNYNWFYGPESQVLFDKFLKNGAGNWLAAQOIRKRPKDGMPVSTAQHRS 393
Db 218 STT-----SKRYE-----GAQITLTNVNDLLTSSGNDILATIQCPQGDGYL-----WP 261
Qy 394 TLHTEYIWDYDQ-----LTPQPPA-----DYGTAKIHTFPNNGVVTYGA 432
Db 262 VIQNGKLTYDERKEKPLGVKSIQLLTNGFPVNLNSMLTGTYYNDTRLSANKELLLYKA 321
Qy 433 GLP-----NTQNTNFTVSFKSGKLGGRAYVDIVHFQPSWIDGWSRFPNGHEHPDQNS 484
Db 322 KIPVASKFFNVSAPTSNSVYLYNAFFAEIEKIYK-----DQYEKQKD 362
Qy 485 FTFAPNGQVFSVSEALYGPKLSH-----LNNVLVAPSPSSQCNKPWE----- 526
Db 363 FTFDHNAITKINGGLL-----LNHKWDSNNALSY--BDNRYPFKLMWDEGFKIRASFNFKL 417
Qy 527 -----GOLGBCAOWLKWLTGTEV--GDAAGBIITASQHGEMVFSV-GEAVS----- 568

Db 418 CGTGAQKGGSLTENAEQKKPLADAVVTGQVAGEVST-----VFESIGRNVSQGFNLNN 471
Qy 569 -----AYSSAMRLKSVYRALLLLNSQTL-----LVVDHLIERQEDSPINSVSAP 611
Db 472 LGLWSFNMGFAKTKAEQIKBATK--IGLNNITMPGTLNLTFLNH-----NAP 517
Qy 612 FHNLDIDFKYIPYKPMN--RYN-GAMMDVDAHYKMFWDHGHGNSPMASIOEAEQAAB-F 667
Db 518 MYDL--KYLPSYKETIRQNLGELID-----NTPLSLGKVRGIVDIL 559
Qy 668 KKRQTQFVNVTFQMESTITRIAYFYGPYINVSRCRFDSSNPNLQIQLSNVNTHEVWSI 727
Db 560 YPGWTKEVDATCYNMSNI-----WLNTRGFENFI---APCLPLELEVEN-----KL 601
Qy 728 VTDYHNL-----KTRFNLYLFGG-----FASVADQ-----GQIT- 756
Db 602 KADLINLDIFTNKPACT-FSPGIGIPAIACCYHFDITDRPAKFVDDDELKVDFDELIGQTTY 660
Qy 757 RFGLGTQAIKPVHRDRII-----FPFGFK--FNIAVG-----LILCISL- 794
Db 661 RUGTVIRKDKKPLYWDSPEALPPNTPLSYEPAVINIASGKREDVNRVKYPVIDTIDIK 720
Qy 795 -----VILTFQWRFYLSFRKLMRWILILVIALWFIELDVMTSCQ-----PICAKWT 842
Db 721 GLGKCDIKLTAPSGFYSPDARKD-----EFKNVYEAIFESNGKYTDDMSQWS 768
Qy 843 RTEAGSKSLSEGHMDLPDVITSLPGSCAEILKQLPFNSSDPLVIRVPTAIDPIE 902
Db 769 NYINFSTLDEINSVGTLSYPK-----PPVETDFLKNTKNSDDGVNASLQT----- 815
Qy 903 TELEIDSFPVDAECWKSVDIIRSGHFRLLRGWLQSLVQDTKLHLQNLHLHPEPNRGKLAQVPA 962
Db 816 -----INETLYKGD-----NSLLSATQYKRPFRFYFEPDSSKLANLFG 854
Qy 963 MNKDKRKFRRRESLPEQRQSMKGAFDRDAEYIRALRRHLVYYPSPARVPLSSSGSWTLK 1022
Db 855 Y-----GSKCN 1082
Qy 1077 G---GSKCN 1082
Db 918 NWIQKGTWN 926

RESULT 10
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 09-Jul-2004
C:Accession: JC7350
R:Uchimura, K.; Pasakany, F.; Kadomatsu, K.; Matsuoka, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7350
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: UNIPROT:O99NB0; UNIPROT:O9EP78; DDBJ:AB040710
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylatic
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 1.7%; Score 113.5; DB 2; Length 484;
Best Local Similarity 18.8%; Pred. No. 4.5;
Matches 81; Conservative 72; Mismatches 171; Indels 107; Gaps 18;

Qy 848 GSKSLSEGHM--DLPDVITSLPGSCAEILKQLPFNSSDPLVIRVPTAIDPIE 905
Db 83 GSPGNLSAAGEAVTQEKQHIYVHATWRTGSSGLGELFNQHPDVFYLYBPMWHL----- 135

Qy 906 EIDSFVDACBWKVSDIRSGHFRLLRGWLQSLVQD---TKLHLQNIHLHEPNRGKLAQYFA 962
Db 136 -----WQA--LYPGDAESIQGALRDWLSLFRCDPSVLRLYAQPGDGPGERAPDSA 183
Qy 963 MNKDKKKPKRR-----ESLPEQSQMKGAFRDAB-----YIRALRRHLVY 1005
Db 184 -NUTTAMLFWRWTKVJCSPLCPAARARADVGLVEDKACESTCPVPVSURALEABCRKY 242
Qy 1006 PSA--RPVLSLSSGWSLKLHFFOEVL-----GASMRALYIVRDPRAIYSLMYLSNKP 1056
Db 243 PVVVIKDVRLDLG-----VLVPLLRDPLGLKLVQVLPFRDPA-VHNSRLKSRQ 290
Qy 1057 SL-----YSLKNVPBHLAKLFIEGKGKCNKNSGYAFEBYBJRKLKSKSNAVSL 1109
Db 291 GLLRESIQVLRTRQGRDGHFRVLLAHG-----VDARPGQARALPSAPRADFFLTLSALEVI 346
Qy 1110 SHLWLANAALRINTLPT-----SYQLVKFEDIVHFPQKTERIFAFGLIPLS 1160
Db 347 CEAWL-----RDLFTRGAPAWLRRYRLRYEDLVQOQALRLLRPSGL-RT 395
Qy 1161 PASINQILFATSTNLFYLPVEGEISPTN-----TNVWKNLPDEIKLIENICWTL 1211
Db 396 LAALDAFAPNTRGSAY-----GADRPFLSARDAREAVHVRRLSQEVQVETACAPA 451
Qy 1212 MDRLGYPKFMD 1222
Db 452 MRLLAYPRSGD 462

RESULT 11
S46715
hypothetical protein YHR099w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein H9332.1
C;Species: Saccharomyces cerevisiae
C;Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 09-Jul-2004
C;Accession: S46715
R;Vaadin, M.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 9332.
A;Reference number: S46715
A;Molecule type: DNA
A;Residues: 1-3744 <VAU>
A;Cross-references: UNIPROT:P38811; EMBL:U00060; NID:9487928; PID:94879
C;Genetics:
A;Gene: SGD:TRAI
A;Cross-references: SGD:S0001141; MIPS:YHR099w
A;Map position: 8R

Query Match 1.7%; Score 113.5; DB 2; Length 3744;
Best Local Similarity 18.6%; Pred. No. 1.1e+02;
Matches 256; Conservative 196; Mismatches 482; Indels 439; Gaps 66;

Qy 26 MFAPSTFEE-----SVSNSEWAVFTDDIDQPKQVQDFR-----NOKLKXSMHLPSL 75
Db 218 MFSFKILSECPITWTVLYSSY-----KQLTSTSLPEFTPLIMNLNIQIKQOQ----- 265
Qy 76 YFDAGEIQMRQKSRASHLHLFRAIRAVTMLSNTPTY---YLPPEKHADE----- 123
Db 266 --EARE---QAESRGEH-----FTSISTEINRPAYCDFILAQIKATSLAYVFIRGY 313
Qy 124 AAKNNEYGNLPLALYCLL-CPEDKVAPEFVLEYMDRV----GYK-----DNLVEN 172
Db 314 APEFLQDYVNFVPLDIIRLQDCSELSARKELHATHRLSTNYKKLFLPKLDYLF- 372
Qy 173 APGEVPIGHSLT-----GFATFDLYNLNDHRRQKYLEKIWIITEEMEYS- 221
Db 373 ---ERILIGNFTMHETLRPLAYSTVADFHNIRSELQLSIENTIKIYTGILLDESIAL 429
Qy 222 KVRSGKQLLHNHOATNMIALLTGALVTGVQDKGKANIWQAVVDVMEKTY---FLNHI 278
Db 430 TVQIMSAKLLN-----LVERILKLGKNPQOAPRAKLLMIIDSYNRPKTLNRQ 481

Qy 279 VDSLDEGVAYGYSYAKSVTQYVFLAQRHFNINLNNWLNKMHFWFYATLLPGFQRTVG 338
Db 482 YDTIMK---YIGRYETHKKEAEKJK-----NSIQDN----- 510
Qy 339 IADSNYNWFVGPESQVFLDKFLKNGAGNWLAAQIRKHRPKDQPMVPSAQRMSTLHTE 398
Db 511 -----DKESE--EFMRK-VLEPSDDDLHPQPKKEDINSP----- 543
Qy 399 YIWYDQLTPQPPADYGTAKIHTFPNMGVVYTGAGLPNTQTN-----TFVSFKS 447
Db 544 ----DVENTESDKVKNDEMVEDIKNVAPILL---LP-TPTNDPIKDAFYLYRLTMSFLK 595
Qy 448 GKUGGRAVDIVHFPQYSWIDGWRSPNPGHEHPDQNSFTFAPNGQVPSVSEALYCPKLSHL 507
Db 596 ----TIHDLKVFNP-----PPNEYTVANPKLWASVSRVSEY----- 629
Qy 508 NNVLVAFAPSSQCNKPEWQGLGCAQWLAKWTGEEVGDAAAGEIITASQHEGMVSVSEAV 567
Db 630 -EVIVF-----KDLFHECIIGLKFKHNEKLSPE--TTKKHFD-ISMPSLPV 673
Qy 568 SAYSSAMRLKSVYRALLLNSQTLVVDHIERQ-----EDSPINSVSFAFFHNLDI- 617
Db 674 SATKDARELMD-YLAFMFMQMDNATFNEIIIEQELPFVYVERMLEDSGLLHVAQSFLTSEIT 732
Qy 618 --DPKIYIPKPMRYNGAMMDVMDAHYKMFWFHDHGNSPMASIOEAQAAEFKRWTFQV 675
Db 733 SPNFAGILLRFL---KGKLDLG----- 752
Qy 676 NYTFQMESTITR---TAYVFYGPYINVSCEFIDSSNPGLOISLVNNTTEHVHVSIVTDYH 732
Db 753 NVDFTNSVNLIRLFKLSFMSVNLFPNINEVVLPHLNDLILNSLUKYSTTAEPLVY--FY 810
Qy 733 NLKTRFNLYLGFQGF----ASVADQGITRFLGLGTQAIKVPVRHDRIIFPFGFKENIAVGL 788
Db 811 LIRTLFRSIGGFENLYRSIKPIQLVQLSLNQMLTARLPHERELY-----V 859
Qy 789 ILCISLVILTFQWRFYLSFRKLMRWILILVIAL-WFIEL-----LDVNSTCSQPICAKWR 843
Db 860 ELGITVPVRLSVLAPLPF--LMK--PLVFALQOYVPLVSQGLRTLELCIDNLNTASYFD 914
Qy 844 TEAGSKKSLSSGHHMDLPDVVITSLPGSGAELIKOLFNFSSDFLYIRVPTAIDIPE- 902
Db 915 PTEPVIDDVSKALFNLLQPPFNHAIHNVVRILGKLGGRNQFL--KPPT--DLTEK 969
Qy 903 TELEIDSFVDACBWKVSDIRSGHFRLLRGWLQSLVQDTKLHLQNIHLHE----- 951
Db 970 TELDIDAID---FKINGMPEDVPLSVTPGQISALNLIQSKSDIHYRKSAYKYLTCVLL 1026
Qy 952 -----PNRGKLAQYFAMNKDK-----XKKFKRRRESL-----P 978
Db 1027 LMTKSSAEFTPTNTELLKTAVNSIKLERIGIEKNFDLEPTVNRKDYSNQENLFLRLLESV 1086
Qy 979 EQRSQMGAFDRDAEYIRALRRHLVYPSARPVLSL--SGSWTLKLHFPQEVLGASM-- 1034
Db 1087 FYATSIKELXDDAMDLLNLLDHFCLLQVNTTLNKRKNYNGTFNIDLPNPFMDSSLIL 1146
Qy 1035 -----RALYI--VRDPRAIYSLMYLSNKPSSLY-----SLKNVPEHLAKLF----- 1072
Db 1147 DAIPLFALSYIPIPEVREVGVLAYKRIYEKSLIYGEELALSHSFPIE-LAKOFIHLCYDET 1205
Qy 1073 --KIEG-----KGKCNLSNGYAFYEPURKELSKSKSNVALLSH 1111
Db 1206 YNKRGGVGLIKVLIIDNVKSSSVFLKKYQYNLANGLLF-----VLKDTQSEAPS----- 1254
Qy 1112 LWLANTAALRINTDLLPTSQYLVKVEDIVHFPQKTERIFAFGLIPLSPASLNQILPAT 1171
Db 1255 ---AITDSAEKLIIDLLSITFADVKEED-----LG-----NKLVENT 1288
Qy 1172 STNLFYLPYGEISPTNTNVMKONLPDEIKLIENICWTLMDRLGYD--KFMD 1222
Db 1289 LTDIVC-----EUSNANPKV--RNACOKSLHTISNL-----TGIFIVKUMD 1327

RESULT 12

S78398
hypothetical protein 2216 - beechdrops plastid
C:Species: plastid Epifagus virginiana (beechdrops)
C:Date: 25-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: S78398; S78404
R:Wolfe, K.H.; Morden, C.W.; Ems, S.C.; Palmer, J.D.
J. Mol. Evol. 35, 304-317, 1992
A:Title: Rapid evolution of the plastid translational apparatus in a nonphotosynthetic
A:Reference number: S78378; MUID:93021155; PMID:1404416
A:Accession: S78398
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2216 <MOL>
A:Cross-references: UNIPROT:P30072; EMBL:M81884; NID:g336917; PID:g336938
A:Genetics: G1
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
A:Accession: S78404
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2216 <MOL>
A:Cross-references: EMBL:M81884; NID:g336917; PID:g336938
A:Genetics: G2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
C:Genetics: <G1>
A:Gene: 2216_a
A:Genome: plastid
C:Genetics: <G2>
A:Gene: 2216_b
A:Genome: plastid
C:Keywords: plastid

Query Match 1.7%; Score 113; DB 2; Length 2216;
Best Local Similarity 17.6%; Pred.No. 52;
Matches 234; Conservative 187; Mismatches 423; Indels 488; Gaps 66;

Qy 42 WAVFTDDIDQTKQVQDFRPNOKLKKMLHPS-LYFDAGEIOAMRQKGRASHLH--LF- 97
Db 408 YSVISNEDSEYHTLINO--REIKPLKSIIPDPSPFLQTEATEIESDQLOKRPSCYSTLT 465
Qy 98 -----RAIRSAVTNLSNPTYYLPKHAADFPAKWEIY-GNNLPLAL--YCLLCP 146
Db 466 EHEKOMINHLPEIEEFIGNPRLV---HSLFSDRWSELHSGNPTSTRDRHKLKK 521
Qy 147 EDKVAR-----EPV-----LEYMDRMVGK-----DLVENAPGDRVPIGHSL 184
Db 522 QDDLSFVPSRRSENKELVNLIKITYLNTVSIHPISSDPGCGVLKDEP--DMDSNKKI 579
Qy 185 TGF-ATAFDFLYNLLDNHRR-----QKYLEKIWV-----ITEEMEYYSKVR 225
Db 580 SVFNKNTFIYLFPHLPHDNVRVGYTLNLHDDFELEERFOEKADLFTLSITEPDLVYHKGS 639
Qy 226 WKKOLLHNHQAQNMIALLTGALVTGVQDKGSKANIKQAVVDVMEKT-----MFL 275
Db 640 FS---IYMDOKQKQVAVPASNIMEAVNQSR---FIRNMKIQYSTYGYIRNVLHFRFLM 692
Qy 276 NHIVDGLSDGVAY-----GSTAKSVTYVFLAQRP-NINLNDNWLKWHFWFYA 327
Db 693 NR-----SDHNLEYIKDQIGKDTLNHRHTIIKYMINOHLNPKKSNKWFNPILF- 743
Qy 328 TLLPGFQRTVGIADSNYNWFGPESQVLFDKFKLXGAGNWLAAQIRKHRPKDGMVPS 387
Db 744 -----FSRTSERNR-----PDAY-----RYKRSNGSNFLE----- 771
Qy 388 TAQRMSTLHTEYIWDPLTPQPADYGTAKIHFTFNPWGVVTVGAGLNTQNTWTFVSGK 447
Db 772 -----HLEH-----FVSEOK 781
Qy 448 KGLGGRAYDVIHVFQPSWIDGWRSPNPGHEHPDQNSFTFAPNGQVVFSEAL----- 501
Db 782 SHPKFKIVFLIIRFNQVS-ID-WSAFIDTKDLKSLPLRF-FLSKLLPLFLSNLPPFFCVSG 838
Qy 502 PKLSHLNNVLVF-APSPSQCNKPWEGQLGECQWLKWTGEEVGDAAGEIITASQHGEMV 560

Db 839 NIPHIRSEIYIELKDPNDQL-----CNQFL-----EPIDL--KIVHLKKRKPP 881
Qy 561 FVSGAVSAVSAMBLKSVYRALLLNSQTLVVDHIER-QEDS--PINSVSAFHNLDI 617
Db 882 -----LVHGTSTRKUK-----LITGTRPPLFNKIPRCMIDSFHTINNRKSDNTD- 928
Qy 618 DPKYIPYKPMNRNGAMDMVDADAHYKMFWDHGHGNSPMASIQEAQAAEFKKRWTFQVFN- 676
Db 929 --SYLSMIFHNK-----DNMLNLVKPF-----HRSLSISYFYKANR-----LQFLNN 968
Qy 677 -----VTFQWESTITRIAYVFGPYINVSSCRPIBSSNPGLOISLANNVNEHV- 724
Db 969 PHNFCYCNTRLPPFYVEKAHINYFTYQFLNLFIR-----NKIFSLCVDKKKCHAF 1021
Qy 725 -----VSIV-----TDYHNKTRFNY-----LGFGGFASVADQ 752
Db 1022 WGRDRTIPBIESQVSKIFIPKNFPOSQDETYNLSQPHFPSPRYDPFVFLIANIYGTPLTE 1081
Qy 753 GQITRFGLTQAIKVPVRHRIIPFG-----PKFNIAVGLI----- 789
Db 1082 GQIVNLG---RTYCOPLSDMLNDSQGNFQYLNFSNMGLIHTPCSDKYLPSKKRKR 1138
Qy 790 -LCIS-----LVILTQWRFYLSFRKLMRWILILVIALWFI----- 824
Db 1139 SLCINKYKCEKGQMYRTFORK--VAFSTLSKNWLFQTYMPWFLTSAGYKYNLIFLDTF 1196
Qy 825 -ELLDVWSTCSQPICAKWTRTAEQSKLSSEGHMDLPDQVITSLPSCGAEILKQLFF 883
Db 1197 SELLSILSS-----SKFVS-----IFNNIMHSG----- 1221
Qy 884 NSSDFLYIRVPTAYIDIPETELEIDSFVDACEWKVDSIRSGHFRLLRGLWQLSQVDTKLH 943
Db 1222 -----ISWRIINKKCLPQWNL-----ISEISSKCLNLLSEETIRQNNEP 1264
Qy 944 LQNIHLHEPN-----RGKLAQYFAMNKDKKKPKKRRESLPQSQRM 984
Db 1265 LIJTHLRSPNVRFLYSILFLLLVGVYLVTRHLLFVSRSASSELOTEFRKRVSLMIPSSMI 1324
Qy 985 KCAFDRAEYIRALRHLVYVPSAPVLSLSSGSWTLKLFHQFVGLGASMRALYIVROPR 1044
Db 1325 E-----LRKLNRYPT-----PASNSFW-----LKNLFIVA--- 1350
Qy 1045 AWIYSMLYNSKPSLSYSLKNVPEHLAKLF-KIEGGKGNLNSGYAFYEPL----- 1094
Db 1351 -----MEQLVYSLEEIRASGGLNLLGPAYGVKISCKNKFYNILIDLIPNPIRI 1401
Qy 1095 -----RKELSKSKSNVALLSHL-----WLANTAALRINTDLLPTSQOLVKFEDIVHP 1144
Db 1402 FSRNMRHLSHTSKEYISLIRKKNVNGWDIIEBSWANSIDDEER---EPLVQFS 1457
Qy 1145 QKTE-RIFAPL 1155
Db 1458 ALTTEKRIYQIL 1469

RESULT 13
S76412
hypothetical protein elr0408 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S76412
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4199 <KAN>

A;Cross-references: UNIPROT:P74440; EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BA01854
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C;Superfamily: Synechocystis hypothetical protein elr0408

Query Match: 1.7%; Score 112.5; DB 2; Length 4199;
 Best Local Similarity 20.9%; Pred. No. 1.5e+02;
 Matches 94; Conservative 42; Mismatches 131; Indels 183; Gaps 25;
 224 RSMGKQLLH-NHQATNMIALLTGALVTGVDGSKANI-----WKQAV 264
 3062 QSMGKLPALNELAVNNTSGNPQIAGL--GGKGGIEYNGSTWVNGPYQGQWRSAL 3119
 265 VDVMEKTMFLNHNHVDGSLDE--GVAYGSYT-----AKSVTQYVFLAQRH 307
 3120 -----TQMAVQWGEDGSPSQIVVGLADGAVIYNTQSGWRTINNFGKSVTQLSVQWQEA 3173
 308 PNIN--NLNHN-----WLKMH--FWFYATLLPGFQRT-----VGIADS 342
 3174 SNPNIVVGLDNSEVQYQSGNSGVMTQFDDGWMVYPVQQL-AVQWTSNDAQPLVVVGLGDD 3232
 343 NVN---WYF---GPESQLVFLDKFKLNGAGNLAQIQRKRPKDGMPVSTAQWSTL 395
 3233 NGNNGSVWYQSGEGGQWTFSG--LPSGAA-----IAQMAVQWNF- 3272
 396 HTEYIWDPPQTPPPADYGTAKIHTFPNMGWVTVYAGLPNTQTNTFVSKGLGGRV 455
 3273 -----SSPNPPNNVNDLKI-----VV-----QADSTVSYNG----- 3301
 456 YDIHFQPSYIDGWR-----SNPGHEHPDQNSFT--PAPNQVVFSEALYQPKLSHLNV 510
 3302 -----NGWTATPAINSSLIQIPTLNAITVQWSANGQPQITVGLGDPYD--NGQ 3347
 511 LVFAPSPS-----SOCKNPEGLGECAG-----WLKWT 539
 3348 LWYLPNPSQSQWELQSGVWYASPTQIDSSWTESLVNPSQTDNLVYVFGSDFNQVNT 3407
 540 GE-----EVGDAAGEIITASQHGEMVVFSG 564
 3408 GTIGDDVMVGSATGESFLAGGDDQILTKG 3437

RESULT 14
 AE1698
 isoleucyl-tRNA synthetase [imported] - Listeria innocua (strain Clip11262)
 C;Species: Listeria innocua
 C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AE1698
 R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A;Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.
 A;Reference number: AB1077; MUID:21537279; PMID:11679669
 A;Accession: AE1698
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-921 <GLA>
 A;Cross-references: UNIPROT:Q92926; GB:AL592022; PIDN:CAC97357.1; PID:g16414641; GSPDB:Q
 A;Experimental source: strain Clip11262
 C;Genetics:
 A;Gene: ileS
 C;Superfamily: isoleucine-tRNA ligase

Query Match 1.7%; Score 111; DB 2; Length 921;
 Best Local Similarity 18.8%; Pred. No. 19;
 Matches 174; Conservative 112; Mismatches 306; Indels 334; Gaps 46;
 117 PPKHADFAAKNNEIYGNLPLALYCLCPEDKVAFFVLEWMDRMVGYKWLVENAP-- 174
 21 PNKEPEWQAKW-----BEEKLYEKIQE---KNAGRKAYILHDGPPY 58

175 -GDEVPIGH-----SLTGFATAPDFLYNLLDNHRR--OKYLEKIWIITEEM 217
 59 ANGELMHGHALNKTIKDIIIVRYKSMAGFSSPY---VPGWDTGHLPIETAIAKGVKREKM 115
 218 -----VEYSKVRSMGKQ-----LLHNHQAATNMIALLTGALVTG 250
 116 SIAEFRKLCABEYANTQVDGORTGPKRLGINDGNDENPIYITLLPEYEAQIKVFGEMAKGY 175
 251 VDKGSKANIWK-----QAVVDVMEKT---MFLNHNHIVDGS--LDEGVAYGSYTAKS 297
 176 IYKGGKPVYSPSESALAEIEYQDKTSASIFVAEKVTDGKGLDGTGVIWIT---T 232
 298 TQYVFLAQRHFNIN-NLDNNWLMKHFWFYA--TLLPGFORTVGIADSNVNFVGPESQL 354
 233 TPMTIPANMIGITVNPDLDDYVVISAGEKYVVAEALLPSLREKLG----- 276
 355 VFLDKFILKNGAGNLAQIQRKRPKDGMPVSPAQWSTLHTEYIWDPPQTPPPADY 414
 277 -FEDATVTVTRGSELDRVVTKHPFFYDRDSLVMNGE-----HATAEA 317
 415 GTAKIHTEFP-----NMGVVTVYGA---GLPNTQTNTFVSKSG 448
 318 GTGAVHTAPGHGEDDPLIGKYYDLEILAPLDDRGVFTTEAPGFGVFDYTKANKVTEKLE 377
 449 KLGGRAYVD-IVHFQPSYIDGWRSPNPGHEHPDQNSFTFAPNQVVFVS-EALYQPKLSH 506
 378 EVGALLKMEFITHSYPHD---WRTKRP-----VIFRATAQWFAIDAFRDLDLAA 424
 507 LNNVLPAPSPSSQCNKPEWEGQCEQAOLKWTGEEVGDAAAGEIITASQHGEMVVFSGEA 566
 425 VKGV-----NWT-----PAMGE---TRLFNMYDRDGDW 449
 567 VSAYSSAMRLKSVYRALLNLSQTLVVDHIEROE---DSPINSVSAFF--HNLIDIDFK 620
 450 VISQRANGVP-----LPIFYAENGELITDETINHISELFREHGSNVWFE 495
 621 -----YIPYKFMN--RYNGAM---MDVMDAHYKMFWDHNGHNSPMASI-----QEAQ 663
 496 RDVKDLLPAGFTHPGSPNGEFTKETIMDV-----WFD-SGSSHQAOLNARPELSRPADL 549
 664 AAEFKKWTQFVNTFQWESITRIAYVFGYPIYVNSCRF-IDSSNPGLOISLVNNT- 721
 550 YMEGSDQYRGWEN-----SSLTTAVAITGEAPYRNVLSHGFDLGE--GRKMSKSLGNTL 602
 722 -----EHVVSIVTD-YHNLKTRFNLYGFGGFASVA 750
 603 LPQKVIKQLGADIVRLVWASVDYQADVRSDEILKQVSEYVRKXINTWRF- 653
 751 DQGITRFGLGTOAI-----VKPVRHDDRIIFPPGFKFNIAVGLILC 791
 654 -LGNINDFNPTTNGSVYENLRVDKYMLIKLNDLVKNVKDSYEAPEFSTIVH-QINNPFCT 711
 792 ISLVILTFQWRFYLSFKLMRWILIL-----VTALMFIELLDVWSTCSQICAKWTRTE 845
 712 VELG-----QFYMDFAKDVVYIEADSHDRRAMQTVFYEAAVTLTKLAPILPHTT--- 762
 846 AEGSKKSLSSEG-----HHMDLPDVVI 867
 763 -EEVWNSLIGEGAESIHLDLPDVKV 787

RESULT 15
 C89874
 autolysin [imported] - Staphylococcus aureus (strain N315)
 C;Species: Staphylococcus aureus
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
 C;Accession: C89874
 R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; ma, A.; Hiratsuka, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
 Lancet 357, 1225-1240, 2001
 A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: C89874
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1248 <KUR>
A;Cross-references: UNIPROT:Q99V41; GB:BA000018; PID:g13700854; PIDN:BAB42150.1; GSPDB:G
A;Experimental source: strain N315
C;Genetics:
A;Gene: atl

Query Match 1.7%; Score 110.5; DB 2; Length 1248;
Best Local Similarity 18.3%; Pred. No. 32;
Matches 245; Conservative 146; Mismatches 462; Indels 487; Gaps 66;
30 STEPEVSNSYSEWAVFTDDIDQKTO--KVOPFRPNQKLKSMLEHP--SLYFDA----- 79
187 TTFSASQPSVAATPKTSIPKYPQVNSINDYIRKNLAKPKIBEDYTSYFPKAYRN 246
80 -----GEIQMRQKRSASHLHLP-----RAIRSAYTVMLS----- 109
247 GVGREPIVVHDTANDRSTINGEISYMKNNYQNAFVHAFVDGRIIETAPTDLYSNGVGA 306
110 --NPTYYLPPKH-----ADFAAKNEIYGNLPLALYCLLPEDKVAFE 153
307 VGNPRFINVEIVHVDYASPARSMNNYADYAATQLOYG-----LKPDS----- 350
154 FVLEYMDRMVGYKDWLVENAPG--DEV-PIGHSI.TGFATAFDELXNLLDNHRRQKYLEKI 210
351 --AEYDNGVTWTHYAVSKYLGTDHADPHGY--LRSHNYSYDQLYDLIN-----EKYLIKM 403
211 WVITEEMYEYSKVRSMGKQLLNH-----QATNMIAL-----TGAL 247
404 -----GKVA PWGTQTFTTPTPSKPTPSKPTGKLTVAANNVGAQIKPTNSGLY 453
248 VTGVDKGSKANIWKQAVVDWNEKTMFLNHI VDGSLDEGVAGSYTAKSVTQVVFIAQRH 307
454 TTVDYDTGKA-----TNEVQKTEAV-----SKTATLGNOKFYLVQDY 490
308 FNTNLLNNLNMKHFYFYATLTP-----GQRTV--GIADSNYN 345
491 NSGNKF--GWNKGDVVINTAKSPVNVNQSYSTKSGTKLTVPWGTSKQVAGSVSGSGNQ 548
346 WFGPESQLVFLDKFIL-----KNGAGNWL--AQOIRKRPKDGPM--VPSTAQRWSTLHTE 398
549 TFRASKQOQI--DKSIYLYGVSNGKSGVSKAVLVDATAKPTPTPIPKPSTPTTNKLTVS 606
399 YIWDPOLTPQPADYGTAKIHFPN--WGVVTVAGLPTQT--NTFVSPKSGKLGRAV 455
607 SL-----NGVAQINAKNGLFTTVYDKTKPTKEVQKTFVAVTKEASLGKNKF 653
456 YDIVHFQPSWIDGWRSFNPGHEHPDONSFTFAPNGQVFVSEALYGPKLSHLANNLVFAP 515
654 YLV-----KDYNSPTLI--GWNKGDVIYNNAKSPVNMQTYTV 690
516 SPSSQC-NKPWEGOLGCAQWLKWTGEEVGDAAGEIITASQHGEMVFV-----SCEAV 567
691 KPGTKLYSVFW-GTYKQEGAVSGTGNQTFKATKQ-----QQIDKSIYLFCTVNGKSGWVS 745
568 SAY-----SSAMRLKSVYRALLNLSQTLVVDHIERQEDSPINSVSAFFHNLDIDF 619
746 KAVLAVPAAPKKA VAOPKTA KAVYTVTKPTQTQTVSKIAQVK--PNNT----- 791
620 KYIPYKEMRYNGAMMDVDADHYKMFWDHGHGSPMASIQEAQAAEFKRWTFQFVNVT 679
792 -----GIRASVYE-----KTAKNGAKYADR-----TF 813
680 QMESTITRIAYFYGPYINVSRCRFDSSNPGQLISLNVNNTTEHVSVISVTDYHNLKTRFN 739
814 -----YVTKERAH-----GNETYVLLNNTSH--NIPLGWENVKD-LN 847
740 YLGGFGFASVADQOQITRFGLTQAI V KPVVRHRIIPFGFK-----FNIAGVLICI 792
848 VQNLGKEVKTKYKTVNKSNNGLSMV-----PWGTKNQVILTGNNIAQGTFNAT 896

QY 793 SLVILTQWRFYLSFRKLMRWILILVIALWFIELLDVMSTCSQPICAKWTRTEAGSKS 852
DB 897 KQVSVGKDVLYGTINNRTGWNA-----KDLTAPTAVKPTTSAAKDYNVT 942
QY 853 L-----SSEGHMDLP--DVVITSLPGSGAELKQLFNSSDFLYIRVPTAYIDIPELEI 907
DB 943 YVIKNGNGYVYVTPNSDTAKYSLKA-----FNEQPFVAVK----- 977
QY 908 DSFVDACEWKVSDIRSGHFRLLRGLQSLVQDTKLHLQNIHLHPEPNRGKLAQYFAMKDK 967
DB 978 EQVINGQWYVYKLSNGKL-----AWIKS-----TDLAKELI----- 1009
QY 968 KKKFKRRRESLPQRQSMKGAFFRDAEYIRALRRHLVYYP SARPVLSLSSGSWTLKLHFFQ 1027
DB 1010 --KYNQGTMTLNQVAIQAG-----LQYKPVQVRV-----PGKWTDA--NFN 1047
QY 1028 EVLGASMRALYIVRDPRAWIYMYSLYNSKPSLYSLKNVPEHLAKLPKIEGGKGCNML-NSG 1086
DB 1048 DVKHA-MDTKRLAQDP-ALKYQFLRLDQPNISIDKINQFL-----KGGKGVLENQ 1096
QY 1087 YAFEVEPLRKELSKSKSNVSLLSHLWLANTAAALRINTDLLPT---SYQLVKFEDIVHF 1143
DB 1097 AAP-----NKAQMYGINEVYLI SHA-----LLETNGTSQLAGADVNV- 1136
QY 1144 POKTTERIPAFILGIPLSPASLNQILFATSTNLFYLPYEGEISPTNTNVMKQNLPRDEIKL 1203
DB 1137 -----NKVV--TNSNTKYHNVFGLAAVDNDPL-----REGIKY 1167
QY 1204 IENICWTLMDR--LGYPKFM 1221
DB 1168 AKQAGWDTVSKAIVGGAKEFI 1187

Search completed: June 23, 2005, 08:50:24
Job time : 34.7171 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2005, 08:33:13 ; Search time 134.102 Seconds
(without alignments)
4666.302 Million cell updates/sec

Title: US-10-697-828-15

Perfect score: 6532

Sequence: 1 MPKGGAPPWIMALMFTGHL.....LIENICWIMDLRLGPKFMD 1222

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6527	99.9	1222	2 Q8IZU8	Q8IZU8 homo sapien
2	4186	64.1	784	2 Q5P523	Q5P523 homo sapien
3	4037	61.8	755	2 Q3HAK5	Q3HAK5 homo sapien
4	2983.5	45.7	597	2 Q8BZP3	Q8BZP3 mus musculus
5	2597	39.8	480	2 Q9H951	Q9H951 homo sapien
6	1757.5	26.9	958	2 Q8BL14	Q8BL14 m mus muscu
7	1753.5	26.8	958	2 Q9UL01	Q9UL01 homo sapien
8	1331	20.4	314	2 Q9HAD7	Q9HAD7 homo sapien
9	1242	19.0	230	2 Q6J9G8	Q6J9G8 gorilla gor
10	1194	18.3	222	2 Q6J9G9	Q6J9G9 pan troglod
11	251.5	3.9	680	2 Q8CJW4	Q8CJW4 streptomyce
12	192	2.9	776	2 Q8UBJ1	Q8UBJ1 agrobacteri
13	150	2.3	411	2 Q9EQC0	Q9EQC0 mus musculo
14	147	2.3	411	2 Q43916	Q43916 homo sapien
15	145	2.2	2504	2 Q85160	Q85160 photorhabdu
16	143.5	2.2	761	2 Q9KWR4	Q9KWR4 spingomona
17	134	2.1	483	2 Q794G9	Q794G9 mus musculo
18	134	2.1	530	2 Q88276	Q88276 mus musculo
19	134	2.1	530	2 Q80WV3	Q80WV3 mus musculo
20	133	2.0	483	2 Q8UED5	Q8UED5 homo sapien
21	133	2.0	530	2 Q7Y4C5	Q7Y4C5 homo sapien
22	132.5	2.0	1337	2 Q8A3U4	Q8A3U4 bacteroides
23	129	2.0	304	2 Q677T4	Q677T4 lymphocyeti
24	128.5	2.0	532	2 Q644R5	Q644R5 erwinia car
25	128	2.0	486	2 Q9VMC3	Q9VMC3 drosophila
26	127	1.9	441	2 Q93403	Q93403 torpedo cal
27	127	1.9	1456	1 MANR_HUMAN	P22897 homo sapien
28	127	1.9	2958	2 Q7RNZ3	Q7RNZ3 plasmodium
29	126.5	1.9	420	2 Q6DBY9	Q6DBY9 brachydanio
30	126	1.9	388	2 Q9WUE5	Q9WUE5 mus musculo
31	126	1.9	2019	2 Q6BEW0	Q6BEW0 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q8IZU8	PRELIMINARY;	PRT;	1222 AA.
AC	Q8IZU8;			
DT	01-MAR-2003 (Tremblrel. 23, Created)			
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)			
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)			
DE	NCAG1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Goossens D., Del-Pavero J., Van Broeckhoven C.;			
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF480435; AAN32895.1; -			
DR	GO; GO:0008146; F:sulfotransferase activity; IEA.			
DR	InterPro; IPR000863; Sulfotransferase.			
DR	Pfam; PF00685; Sulfotransfer 1; 1.			
SQ	SEQUENCE 1222 AA; 140272 MW; 156BE397B4E99AE9 CRC64;			

Query Match	99.9%;	Score	6527;	DB	2;	Length	1222;
Best Local Similarity	99.9%;	Pred. No.	0;				
Matches	1221;	Conservative	0;	Mismatches	1;	Indels	0;
Gaps	0;						
Qy	1	MPKGGAPPWIMALMFTGHLFLALLMFAFTFEESVSNYSEWAVFTDIDOFKTKQVDF	60				
Db	1	MPKGGAPPWIMALMFTGHLFLALLMFAFTFEESVSNYSEWAVFTDIDOFKTKQVDF	60				
Qy	61	RPNOKLKSMHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKH	120				
Db	61	RPNOKLKSMHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTMLSNPTLYLPPPKH	120				
Qy	121	ADFAAKWNEIYNNLPPALYCLLPEDKVAFEVLEMDRMVGYKDWLVENAFGEVPI	180				
Db	121	ADFAAKWNEIYNNLPPALYCLLPEDKVAFEVLEMDRMVGYKDWLVENAFGEVPI	180				
Qy	181	GHSITGATAPDFLYNLLDNHRRQKLEKIWIITEEMVEYSKVRSGKQLLNHQAQNI	240				
Db	181	GHSITGATAPDFLYNLLDNHRRQKLEKIWIITEEMVEYSKVRSGKQLLNHQAQNI	240				
Qy	241	ALLTGLVTVGDKSGKANIWKQAVVDVMEKTMFLNHNIVDGLDEGVAYGSYTAQSVQY	300				
Db	241	ALLTGLVTVGDKSGKANIWKQAVVDVMEKTMFLNHNIVDGLDEGVAYGSYTAQSVQY	300				
Qy	301	VFLAQRHFNINLNNLKMHPFYATLLPGPORTVGIADSNYNWFFGPSQLVFLDKF	360				
Db	301	VFLAQRHFNINLNNLKMHPFYATLLPGPORTVGIADSNYNWFFGPSQLVFLDKF	360				
Qy	361	ILKNAGNWLAAQIRKRPKDGPMVPSAQRMSTLHTEYIWDPOLTPQPPADYGTAKIH	420				
Db	361	ILKNAGNWLAAQIRKRPKDGPMVPSAQRMSTLHTEYIWDPOLTPQPPADYGTAKIH	420				

```
Qy 421 TFFNWGVVYTGAGLPNTQTNTFVSFKSGKLGGRVAVDVIHFQPSYMTWIDGRSNPQGEHP 480
Db 421 TFFNWGVVYTGAGLPNTQTNTFVSFKSGKLGGRVAVDVIHFQPSYMTWIDGRSNPQGEHP 480
Qy 481 DQNSFTFAPNGQVFVSALYGPGLSHLNNVLFAPSPSSQCNKPWEGQGECAQWLKWTG 540
Db 481 DQNSFTFAPNGQVFVSALYGPGLSHLNNVLFAPSPSSQCNKPWEGQGECAQWLKWTG 540
Qy 541 EBYGDAGEIITASOGEVMFVSGEAVSSAMRLKSVYRALLLNSQTLLVVDHIERQ 600
Db 541 EBYGDAGEIITASOGEVMFVSGEAVSSAMRLKSVYRALLLNSQTLLVVDHIERQ 600
Qy 601 EDSPINSVSAFFNNLDIDFKYIPKPMNRNGAMVDWADHYKQFMDHGHGNSPMASIOE 660
Db 601 EDSPINSVSAFFNNLDIDFKYIPKPMNRNGAMVDWADHYKQFMDHGHGNSPMASIOE 660
Qy 661 AEQAAPFKRWTFQVNVTFQMESITIRIAYVFGPYINVSSCRFDIDSSNPGQLISLVNN 720
Db 661 AEQAAPFKRWTFQVNVTFQMESITIRIAYVFGPYINVSSCRFDIDSSNPGQLISLVNN 720
Qy 721 TEHVVSIVTDYHNLKTRFNVYLGSGFASVADQGITRFGLGTOAIYKPVHRDRIIPFGF 780
Db 721 TEHVVSIVTDYHNLKTRFNVYLGSGFASVADQGITRFGLGTOAIYKPVHRDRIIPFGF 780
Qy 781 KFNIAVGLILCISIVILTFQWRVFLSFRKLMRWILIVIALWFIELLDVNSTCSQPICAK 840
Db 781 KFNIAVGLILCISIVILTFQWRVFLSFRKLMRWILIVIALWFIELLDVNSTCSQPICAK 840
Qy 841 WTRTEAGSKKSLSSSEGHMDLPDVVITSLPGSGABILKQLFPNSSDFLYIRVPTAYIDI 900
Db 841 WTRTEAGSKKSLSSSEGHMDLPDVVITSLPGSGABILKQLFPNSSDFLYIRVPTAYIDI 900
Qy 901 PETELEIDSVDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHFNHGRKLAQY 960
Db 901 PETELEIDSVDACEWKVSDIRSGHFRLLRGWLOSLVQDTKLHLQNIHLHFNHGRKLAQY 960
Qy 961 FANNDKRRFKKRESLPEQSRQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGSWT 1020
Db 961 FANNDKRRFKKRESLPEQSRQMGAFDRDAEYIRALRRHLVYPSARPVLSLSSGSWT 1020
Qy 1021 LKLHFFQEVLGASMRALYIYVRDPRAMTYSMLYNSKPSLSYLNKVPHEHLAKLPIEGGK 1080
Db 1021 LKLHFFQEVLGASMRALYIYVRDPRAMTYSMLYNSKPSLSYLNKVPHEHLAKLPIEGGK 1080
Qy 1081 CNLNSGYAFYEYELRKELSKSNVSLSHLWLANAALRINTDILLPTSQLVKFEDI 1140
Db 1081 CNLNSGYAFYEYELRKELSKSNVSLSHLWLANAALRINTDILLPTSQLVKFEDI 1140
Qy 1141 VHPFQKTERIFAPLGIPLSPASLNOILPATSTNLFLVLYEGBISPTNTNWKQLPRDE 1200
Db 1141 VHPFQKTERIFAPLGIPLSPASLNOILPATSTNLFLVLYEGBISPTNTNWKQLPRDE 1200
Qy 1201 IKLIENICWTLMRLGYPKFMD 1222
Db 1201 IKLIENICWTLMRLGYPKFMD 1222

RESULT 2
Q6P5Z3 PRELIMINARY; PRT; 784 AA.
AC Q6P5Z3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE C18orf4 protein (Fragment).
GN Name=C18orf4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
```

```
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Fellngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC062557; AAH62557.1; -.
FT NON_TER 784
SQ SEQUENCE 784 AA; 89666 MW; 2A86FFC4DDCB846C CRC64;

Query Match 64.1%; Score 4186; DB 2; Length 784;
Best Local Similarity 99.9%; Pred. No. 3.8e-295;
Matches 777; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 11 MALMFTGHLFLALLMFAFSTFEESVSNYSSEWAVFTDDIDQFKTKQVQDFRNQKLKSM 70
Db 1 MALMFTGHLFLALLMFAFSTFEESVSNYSSEWAVFTDDIDQFKTKQVQDFRNQKLKSM 60
Qy 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADFAAKWNEI 130
Db 61 LHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKHADFAAKWNEI 120
Qy 131 YGNLPPALYCLLCPEDKVAPEFVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGATA 190
Db 121 YGNLPPALYCLLCPEDKVAPEFVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGATA 180
Qy 191 FDFLYNLDNHRROKYLEKIWIITEEMEYYSKVRSMGKQLLHNHQAATNMIALLTGALVTG 250
Db 181 FDFLYNLDNHRROKYLEKIWIITEEMEYYSKVRSMGKQLLHNHQAATNMIALLTGALVTG 240
Qy 251 VDKGSKANIWQAVVDVMEKTMFLNHIYDGLDEGVAYGSYTAKSVTQYVFLAQRHFN 310
Db 241 VDKGSKANIWQAVVDVMEKTMFLNHIYDGLDEGVAYGSYTAKSVTQYVFLAQRHFN 300
Qy 311 NNLDNNLNMHFWFYATLLPGFQRTVGIADSNYNWFGPESQLVFLDKFILKNGAGNWL 370
Db 301 NNLDNNLNMHFWFYATLLPGFQRTVGIADSNYNWFGPESQLVFLDKFILKNGAGNWL 360
Qy 371 AQQIRKRPKDGMPVSTPAQRWSTLHTEYIWDPOLTPQPPADYGTAKIHFTFPNMGVVTY 430
Db 361 AQQIRKRPKDGMPVSTPAQRWSTLHTEYIWDPOLTPQPPADYGTAKIHFTFPNMGVVTY 420
Qy 431 GAGLNTQNTFVSFKSGKLGGRVAVDVIHFQPSYMTWIDGRSNPQGEHPDQNSFTFAPN 490
Db 421 GAGLNTQNTFVSFKSGKLGGRVAVDVIHFQPSYMTWIDGRSNPQGEHPDQNSFTFAPN 480
Qy 491 GOVFSEALYGPGLSHLNNVLFAPSPSSQCNKPWEGQGECAQWLKWTGGEVGAAGEI 550
Db 481 GOVFSEALYGPGLSHLNNVLFAPSPSSQCNKPWEGQGECAQWLKWTGGEVGAAGEI 540
Qy 551 ITASQHGEMVFSVSGEAVSAYSSAMRLKSVYRALLLNSQTLLVVDHIERQEDSPINSVA 610
Db 541 ITASQHGEMVFSVSGEAVSAYSSAMRLKSVYRALLLNSQTLLVVDHIERQEDSPINSVA 600
```



```
QY 611 FFHNLIDDFKYIPKFMNRYNGAMMDVMDAHYKMFWDHGNSPMASIQAEQAEEFKR 670
DB 601 FFHNLIDDFKYIPKFMNRYNGAMMDVMDAHYKMFWDHGNSPMASIQAEQAEEFKR 660
QY 671 WTQPVNTVTFQESTITRIAYVFGPYINVSRCRFDSSNPGQLSLNVNNTNTHVSVITVD 730
DB 661 WTQPVNTVTFQESTITRIAYVFGPYINVSRCRFDSSNPGQLSLNVNNTNTHVSVITVD 720
QY 731 YHNLKTRFNLVFGGFSVADOGQITRFLGLTQAIKVPVRHRIIFPFGKFNIAVGL 788
DB 721 YHNLKTRFNLVFGGFSVADOGQITRFLGLTQAIKVPVRHRIIFPFGKFNIAVGL 778

RESULT 3
Q9HAK5 PRELIMINARY; PRT; 755 AA.
ID Q9HAK5
AC Q9HAK5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ11477.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Whole embryo;
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki K., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohayashi M., Nishii T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mueschling K., Yuuki H., Oshima A., Sasaki N., Aotaka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki M., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Koniya M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiroo M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsuura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Satoh N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK021539; BAB13840.1; --
SQ SEQUENCE 755 AA; 86162 MW; A8104C69FA4D6250 CRC64;

Query Match 61.8%; Score 4037; DB 2; Length 755;
Best Local Similarity 99.5%; Pred. NO. 2.4e-284;
Matches 750; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 217 MYEYSKVRSGKQLLHNHQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLN 276
DB 1 MYEYSKVRSGKQLLHNHQATNMIALLTGALVTGVDKSGKANIWKQAVVDVMEKTMFLS 60
QY 277 HIVDGSLEGVAYGYSYAKSVTQYVFLAQRFHNNLNNLKKHFWFYATLLPGFQRT 336
```

```
DB 61 HIVDGSLEGVAYGYSYAKSVTQYVFLAQRFHNNLNNLKKHFWFYATLLPGFQRT 120
QY 337 VGTADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRKRPKDGPMVSTPAQRWSTLH 396
DB 121 VGTADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRKRPKDGPMVSTPAQRWSTLH 180
QY 397 TEYIWDYDQPLTPQPPADYGTAKIHTFPNMGVVTVYAGLNPNTQNTFVSPKSGKLGGRVY 456
DB 181 TEYIWDYDQPLTPQPPADYGTAKIHTFPNMGVVTVYAGLNPNTQNTFVSPKSGKLGGRVY 240
QY 457 DIVHFPYSDWIDGWSFNGCHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLFAPS 516
DB 241 DIVHFPYSDWIDGWSFNGCHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLFAPS 300
QY 517 PSSQCNKPWEGQGECAQWLKWTGEEVGDAAGBIIITASQHGEMVVFVSSEAVSAYSSAMRL 576
DB 301 PSSQCNKPWEGQGECAQWLKWTGEEVGDAAGBIIITASQHGEMVVFVSSEAVSAYSSAMRL 360
QY 577 KSVYRALLLNSQTLVVDHIERQEDSPINSVAFPHNLIDIPKYIPKFMNRYNGAMMD 636
DB 361 KSVYRALLLNSQTLVVDHIERQEDSPINSVAFPHNLIDIPKYIPKFMNRYNGAMMD 420
QY 637 VMDAHYKMFWDHGNSPMASIQAEQAEEFKRWTQFVNVTQFQESTITRIAYVFGPY 696
DB 421 VMDAHYKMFWDHGNSPMASIQAEQAEEFKRWTQFVNVTQFQESTITRIAYVFGPY 480
QY 697 INVSSCRFDSSNPGQLSLNVNNTNTHVSVITVDYHNLKTRFNLVFGGFSVADOGQIT 756
DB 481 INVSSCRFDSSNPGQLSLNVNNTNTHVSVITVDYHNLKTRFNLVFGGFSVADOGQIT 540
QY 757 RFLGLTQAIKVPVRHRIIFPFGKFNIAVGLILCISLVILTQWRFLYSPFKLMRWILI 816
DB 541 RFLGLTQAIKVPVRHRIIFPFGKFNIAVGLILCISLVILTQWRFLYSPFKLMRWILI 600
QY 817 LVIALMFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMMDLPDVVITSLPGSGAE 876
DB 601 LVIALMFIELLDVWSTCSQPICAKWTREAEKSKLSSEGHMMDLPDVVITSLPGSGAE 660
QY 877 ILKQLFPNSDFLYIRVPTAYIDIPETELIDSPVDACEKVSIDIRSGHPELLRGWLQSL 936
DB 661 ILKQLFPNSDFLYIRVPTAYIDIPETELIDSPVDACEKVSIDIRSGHPELLRGWLQSL 720
QY 937 VQDTKLHLQNLHLHPNRGKLAQVFAMNKDKKKR 970
DB 721 VQDTKLHLQNLHLHPNRGKLAQVFAMNKDKKKR 754

RESULT 4
Q8BZP3 PRELIMINARY; PRT; 597 AA.
ID Q8BZP3
AC Q8BZP3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus adult male diencephalon cDNA, RIKEN full-length enriched
DE library, clone:9330132E09 product:weakly similar to SQUAMOUS CELL
DE CARCINOMA ANTIGEN RECOGNIZED BY T CELL (Fragment).
GN Name=9330132E09Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
```

RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RA The FANTOM Consortium;
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa K., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Diencephalon;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yaeunishi A., Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK033983; BAC28538.1; -
DR MGI; MGI:242948; 9330132E09Rik.
FT GO; GO:0005615; C:extracellular space; TAS.
FT NON TER 597 597
SQ SEQUENCE 597 AA; 68197 MW; E34F433EB16EDDC CRC64;
Query Match 45.7%; Score 2983.5; DB 2; Length 597;
Best Local Similarity 90.0%; Pred. No. 6.6e+208;
Matches 542; Conservative 32; Mismatches 23; Indels 5; Gaps 1;
QY 11 MALMFTCHLLFLALLMFAFTEESVNSYSEAVFTDDIDQFTKQVDFRPNQKLKSM 70
DB 1 MAFMFTCHLLFLALLMFAFTEESVNSYSEAVFTDDIDQFTKQVDFRPNQKLKSM 70
QY 71 LHPSLYFDAGEIQAMRQKSRASHLHLFRRAISAVTVMLSNPTLYLPPPKHADFAAKWNEI 130
DB 56 LHPNLYFDAGDIQTLKQKSRTHLHFRRAISAVTVMLSNPTLYLPPPKHADFAAKWNEI 115
QY 131 YGNLPLALYCLLCPDKVAFVFLVMDRMVGYKDWLVENAPGVDPVIGHSLTGFATA 190
DB 116 YGNLPLALYCLLCPDKVAFVFLVMDRMVGYKDWLVENAPGVDPVIGHSLTGFATA 175
QY 191 FDFLYNLLDNRHQRKYLEKIWIWTEEMVEYSKVRSGKQLLHQNATNMIALLTGALVTG 250

DB 176 FDFLYNLLGQRKQKYLEKIWIWTEEMVEYSKIRSGKQLLHQNATNMIALLTGALVTG 235
QY 251 VDKGSKANIWKQAVVDVMEKTMFLNLNHIWDGSLDGVAYGTSYAKSVTOYVFLAQHFNI 310
DB 236 VDKGSKANIWKQAVVDVMEKTMFLNLNHIWDGSLDGVAYGTSYAKSVTOYVFLAQHFNI 295
QY 311 NNLDNNLWKQHFYFYATLLPGFQRTVGTADSNYNWFGPESQLVFLDKFLLKNGAGNWL 370
DB 296 NNFDNNLWKQHFYFYATLLPGYQRTVGTADSNYNWFGPESQLVFLDKFLLQNGAGNWL 355
QY 371 AQQIRKRPKQGMVPESTQAQRMSTLHTEYIWDQPLTPPPADYGTAKIHTFPNMGVVTY 430
DB 356 AQQIRKRPKQGMVPESTQAQRMSTLHTEYIWDQPLTPPPADYGTAKIHTFPNMGVVTY 415
QY 431 GAGLPNTQNTTVPFKSGKLGGRVYDVHFPYPSWIDGWRSNFNGCHEHPDQNSFTFAPN 490
DB 416 GCGLPNTQNTTVPFKSGKLGGRVYDVHFPYPSWIDGWRSNFNGCHEHPDQNSFTFAPN 475
QY 491 GQFVSEALYGPKLHNLNVLFPSPSSQCNKPHGQLGCAQNLKWTGEEVGDAAEI 550
DB 476 GQFVSEALYGPKLHNLNVLFPSPSSQCNKPHGQLGCAQNLKWTGEEVGDAAEV 535
QY 551 ITASQHGEMVFSGEAVSAYSSAMRLKSVYRALLNSQTLVVDHIERQEDSPINSVA 610
DB 536 ITAAQHGDMFVSGEAVSAYSSAMRLKSVYRALLNSQTLVVDHIERQETSPINSVA 595
QY 611 FF 612
DB 596 FF 597
RESULT 5
Q9H951 PRELIMINARY; PRT; 480 AA.
AC Q9H951;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ13005.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tani H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Sato H., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe K., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohata O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs";
RL Nat. Genet. 36:40-45(2004).
DR EMBL; AK023087; BAB14387.1; -;
SQ SEQUENCE 480 AA; 54559 MW; AEF246851B83E9DB CRC64;

Query Match 39.8%; Score 2597; DB 2; Length 480;
Best Local Similarity 99.8%; Pred. No. 5.9e-180;
Matches 479; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 320 MHFWFYATLLPGFQRTVGIADSNYNFYGPESQLVFLDKFKLXGAGNWLAAQOIRKHP 379
Db 1 MHFWFYATLLPGFQRTVGIADSNYNFYGPESQLVFLDKFKLXGAGNWLAAQOIRKHP 60
Qy 380 KDGPMVPSTAQRWSTLHTEYIWDYDQLTTPQPDYGTAKIHTFFNKGWVTVYAGLPLNTQT 439
Db 61 KDGPMVPSTAQRWSTLHTEYIWDYDQLTTPQPDYGTAKIHTFFNKGWVTVYAGLPLNTQT 120
Qy 440 NTFVSEKSGKLGRAVYDVIHFQPVSMIDGWRSPNGHEHPDQNSFTFAPNGQVVFSEAL 499
Db 121 NTFVSEKSGKLGRAVYDVIHFQPVSMIDGWRSPNGHEHPDQNSFTFAPNGQVVFSEAL 180
Qy 500 YGPKLSHLNNVLPAPSPSSQCNKRPWEGQCGEQAQWLKWTGEEVGDAAAGEIITASQHGEM 559
Db 181 YGPKLSHLNNVLPAPSPSSQCNKRPWEGQCGEQAQWLKWTGEEVGDAAAGEIITASQHGEM 240
Qy 560 VFVSGEAVSAYSSAMRLKSVYTRALLNSQTLVVDHIEROEDSPINSVSAPFNLDIDF 619
Db 241 VFVSGEAVSAYSSAMRLKSVYTRALLNSQTLVVDHIEROEDSPINSVSAPFNLDIDF 300
Qy 620 KYIPYKFNRRNGAMVDVMDAHYKMFDFRHNHSGSPMASIQAEQAQAEFKKRWTFVNVTF 679
Db 301 KYIPYKFNRRNGAMVDVMDAHYKMFDFRHNHSGSPMASIQAEQAQAEFKKRWTFVNVTF 360
Qy 680 QMESTITRIAYVFYGPYINVSCHFRIDSSNPGQLQSLNNVNTHEVSVITVDYHNLKTRFN 739
Db 361 QMESTITRIAYVFYGPYINVSCHFRIDSSNPGQLQSLNNVNTHEVSVITVDYHNLKTRFN 420
Qy 740 YLFGGFGFASVADQGGITRFLGTOAIVKPVVRHRIIPFPFGKFNIAVGLIICISLVILTF 799
Db 421 YLFGGFGFASVADQGGITRFLGTOAIVKPVVRHRIIPFPFGKFNIAVGLIICISLVILTF 480

RESULT 6
Q8BL14 PRELIMINARY; PRT; 958 AA.
AC Q8BL14;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
DE enriched library, clone: B130024B19 product: SQUAMOUS CELL CARCINOMA
DE ANTIGEN RECOGNIZED BY T CELL homolog (Squamous cell carcinoma antigen
DE recognized by T cells 2).
GN Name=Sart2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Parthenogenote;
RX Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara K., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okamoto N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akhiba S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
RA Klausner R.D., Collins J.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerker A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [8]


```
Db 342 LADQIRNRVVGEGTSGKQRMCTLTETFLWYDGLSKSVPPDPFGTPTLHFFEDWGVVT 401
Qy 430 YGAGLPNTQNTTFSFKSGKLGRAVDYIVHFQPY-SWIDGWRSPNPGHEHPDQNSFTEA 488
Db 402 YGALPAEINRFSFSGKLGGRAYIDIVHRNKYKDWIKGRWFRNAGHEHPDQNSFTEA 461
Qy 489 PNGQVVFSEALYGPGLSHLNNLVFAPSPSCQNKPMWEGQJGE--CAQWLKWTGEEVYGD 546
Db 462 PNGVPFTEALYGPKYTFNNVLMFSPAVSKSCFSPWGVQVTEDCSSKWSKYKHDLAASC 521
Qy 547 AGEIITASQHGEMVVFSGEAVSAVSSAMRLKSVYRALLLNSQTLVVDHIEREDSPIN 606
Db 522 QGRVVAEEKNGGVFIRGEGVGAYNPQLNKNVORNILLHPQLLLVLDQHLGEESPLE 581
Qy 607 SVSAFFHNLDIDFKYIPYKFMNRYNGAMVDVDAHYKMFNFDHNGSPMASIQEAAE 666
Db 582 TAASFFHNVDVFPF---ETVVDGVGHGAFIRQDGLYKNYMDDTGYSEKATFASVTPRG 638
Qy 667 FKRWTFQVNVVTFQWESTITRIAYVFGPYINVSSCRFDIDSSNPGLOQISLNVNTEHWS 726
Db 639 YPYNGTNYVNVVTHLRSPITRAAYLFGPSIDVOS-----FTVHGDSQQLDVF 686
Qy 727 IVTDYHNLKTRFNLY-----GFGGFASV-ADQSGI 755
Db 687 IATSKHAYAT---YLWTGEATGQSAFAQVIADRHKI 719

RESULT 8
Q9HAD7
ID Q9HAD7 PRELIMINARY; PRT; 314 AA.
AC Q9HAD7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Hypothetical protein FLJ11787.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
TI TISSU=whole embryo;
PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohtsuka M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kinata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mubashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki H., Hiroo M., Ohmori Y.,
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigetani K., Senba T.,
RA Matsumura K., Nakajima Y., Miuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isozaki T., Sugano S.;
"Complete sequencing and characterization of 21,243 full-length human
```

```
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
DR EMBL; AK021849; BAB13912.1; -.
FT NON TER 314 314
SQ SEQUENCE 314 AA; 35955 MW; 722B5555DC065538 CRC64;

Query Match
Best Local Similarity 32.4%; Score 1331; DB 2; Length 314;
Matches 313; Conservative 0; Mismatches 1; Indels 652; Gaps 1;

Qy 1 MPKGGAPPMTALMFTGHLFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKQVQDF 60
Db 1 MPKGGAPPMTALMFTGHLFLALLMPAFSTFEESVSNYSEWAVFTDDIDQFKTKQVQDF 60
Qy 61 RPNQKLKSKMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
Db 61 RPNQKLKSKMLHPSLYFDAGEIQAMRQKSRASHLHLFRAIRSAVTVMLSNPTYYLPPPKH 120
Qy 121 ADFAAKWEIYGNLPLALYCLLPEDKVAPEFVLEYMDRMVGYKDWLVENAPGDEVPI 180
Db 121 ADFAAKWEIYGNLPLALYCLLPEDKVAPEFVLEYMDRMVGYKDWLVENAPGDEVPI 180
Qy 181 GHSITGATAFDFLYNLLDNHRRQKYLEKIWIITEEMVEYSKVRSGKQLLHNNHOATMI 240
Db 181 GHSITGATAFDFLYN----- 196
Qy 241 ALLTGALVTGVDKSGSKANIWKQAVVDVMEKTMFLNHLIVDGLDEGVAYGYSYAKSVQY 300
Db 197 ----- 196
Qy 301 VFLAQRHFNINLNDNNLKMHPFYATLLPGFQRTVGIADSNYNWFPSPESQLVFLDKF 360
Db 197 ----- 196
Qy 361 ILKNGAGWLAQIRKRPKDGMPVSTAOBSWLTHTEYIWDYDQTLPPQPADYGTAKIH 420
Db 197 ----- 196
Qy 421 TFPNWGVVTVYAGLPNTQNTTFSFKSGKLGRAVDYIVHFQPYSWIDGWRSPNPGHEHP 480
Db 197 ----- 196
Qy 481 DQNSFTFAPNGQVVFSEALYGPGLSHLNNLVFAPSPSCQNKPMWEGQJGECAQWLKWTG 540
Db 197 ----- 196
Qy 541 BEVGDAAGEIITASQHGEMVVFSGEAVSAVSSAMRLKSVYRALLLNSQTLVVDHIERQ 600
Db 197 ----- 196
Qy 601 EDSPIINSVSAFFHNLDIDFKYIPYKFMNRYNGAMVDVDAHYKMFNFDHNGSPMASIQE 660
Db 197 ----- 196
Qy 661 AEQAAEFKRWTFQVNVVTFQWESTITRIAYVFGPYINVSSCRFDIDSSNPGLOQISLNVN 720
Db 197 ----- 196
Qy 721 TEHVSVIVTDYHNLKTRFNLYGFGGFASVADQSGITRFGLTQAIKVPVRHDIRIIFPGF 780
Db 197 ----- 196
Qy 781 KFNIAVGLILCISIVILTQWRVFLSPKLMRWLILVIALWFTLELDVWSTCSQPICAK 840
Db 197 ----- 196
Qy 841 WTRTEABGSKSLSSSEGHMDLPDVVITSPLPGSGABILLKOLFFNSSDFLYIRVPTAYIDI 900
Db 197 -----CKKSLSSSEGHMDLPDVVITSPLPGSGABILLKOLFFNSSDFLYIRVPTAYIDI 248
Qy 901 PETELEIDSFVDAACEWKVSDIRSGHGFRLRLGWLQSLVQDTKLHLQNLHLHPEPNRGKLAQY 960
Db 249 PETELEIDSFVDAACEWKVSDIRSGHGFRLRLGWLQSLVQDTKLHLQNLHLHPEPNRGKLAQY 308
```

```

QY 961 FAMNKD 966
Db 309 FAMNKD 314

RESULT 9
Q6J9G8 PRELIMINARY; PRT; 230 AA.
AC Q6J9G8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NCAG1 (Fragment).
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9593;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15123584; DOI=10.1101/gr.1891104;
RT Zhang J., Wang X., Podlaha O.;
RT "Testing the chromosomal speciation hypothesis for humans and
RT chimpanzees.";
RL Genome Res. 14:845-851(2004).
DR EMBL: AY561496; AAT45534.1; -.
FT NON_TER 230 230
FT NON_TER 230 230
SQ SEQUENCE 230 AA; 26793 MW; 1387502DB633B3A3 CRC64;

Query Match 19.0%; Score 1242; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 6.3e-82; Indels 0; Gaps 0;
Matches 230; Conservative 0; Mismatches 0;

QY 152 FEFVLEMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 211
Db 1 FEFVLEMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 60

QY 212 VITEEMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSKANIKQAVVDVMEKT 271
Db 61 VITEEMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSKANIKQAVVDVMEKT 120

QY 272 MFLNLHIVDGLSDEGVAYGYSYTAQSVTQYVFLAQHFNINLNNLNNLNNLNNLNNLNNL 331
Db 121 MFLNLHIVDGLSDEGVAYGYSYTAQSVTQYVFLAQHFNINLNNLNNLNNLNNLNNLNNL 180

QY 332 GFQRTVGIADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRKRPKD 381
Db 181 GFQRTVGIADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRKRPKD 230

RESULT 10
Q6J9G9 PRELIMINARY; PRT; 222 AA.
AC Q6J9G9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NCAG1 (Fragment).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=15123584; DOI=10.1101/gr.1891104;
RT Zhang J., Wang X., Podlaha O.;
RT "Testing the chromosomal speciation hypothesis for humans and
RT chimpanzees.";
RL Genome Res. 14:845-851(2004).
DR EMBL: AY561495; AAT45533.1; -.
FT NON_TER 1 1
FT NON_TER 222 222

SQ SEQUENCE 222 AA; 25736 MW; 4C0CCB8E31AF6CA8 CRC64;

Query Match 18.3%; Score 1194; DB 2; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.9e-78; Indels 0; Gaps 0;
Matches 222; Conservative 0; Mismatches 0;

QY 155 VLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 214
Db 1 VLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYNLLDNHRRQKYLEKI 60

QY 215 EMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSKANIKQAVVDVMEKT 274
Db 61 EMEYKSVRSWKGKOLLHNHQTNNMIALLTGALVTGVDKSKANIKQAVVDVMEKT 120

QY 275 LNHIYDGLSDEGVAYGYSYTAQSVTQYVFLAQHFNINLNNLNNLNNLNNLNNLNNL 334
Db 121 LNHIYDGLSDEGVAYGYSYTAQSVTQYVFLAQHFNINLNNLNNLNNLNNLNNLNNL 180

QY 335 RTVGIADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRK 376
Db 181 RTVGIADSNYNWFYGPESQVFLDKFILKNGAGNWLAAQIRK 222

RESULT 11
Q8CJW4 PRELIMINARY; PRT; 680 AA.
AC Q8CJW4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SC03500.
GN ORFNames=SC0134.01c, SC0565.36c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L.D., Oliver K., O'Neil S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL939116; CAD55321.1; -.
DR PIR: T36083; T36083.
KW InterPro: IPR008929; Chondroitin lyase.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 680 AA; 75582 MW; 631B150D802EF81C CRC64;

Query Match 3.9%; Score 251.5; DB 2; Length 680;
Best Local Similarity 22.6%; Pred. No. 3.1e-09;
Matches 125; Conservative 86; Mismatches 249; Indels 93; Gaps 27;

QY 80 GEIQMRQKSRASHLHLFRATRSATVMSLNPTYYL---PPPKHADFAKWNIEYGNL 136
Db 6 GRDLSLRGDLNGTHAAQWRRLHEQC-----DWYRRQNPPPTPEHEASITY---FGPAAA 55

QY 137 PLALYCLLCPEDKVAPEFVLEYMDRMVGYKDWLVENAPGDEVPIGHSLTGFATFDFLYN 196
Db 56 NLALAYRLTQ--RGYLEANRWISVCVAYPHWGRAHMDHDLDAAGLLHGLSLAWSLGE 114

QY 197 LLDNHRQKYLEKIWIITEEMEYKSV---RSWGKOLLHNHQTNNMIALLTGALVTGVDK 253
Db 115 DLEPERREILRAKLELQGERLHSPAEETTGWRWSSAYWQNH---NWIC-WTGIATAGYAL 170

```

```
QY 254 GSKANIWKQAVDVMEKTMFLNHI VDGSLDEGVAYGTYTAKSVTVQVFLAQRH----- 307
DB 171 GRSE--WTKAARANLETVLWLPB--DGSDEGVVYRYGVFWLAHTDLVQREQADLW 226
QY 308 ----FNINNDNNLWVHFYFYATLPGPQRTVGIADSNYNWPFYGPESQLVFLDKFILK 363
DB 227 STGFLRNT--TRW-RLH-----QSAPGPEENIDHGDC-HDRSRGHSVALYVRLASAYQ 276
QY 364 NGAGNMLAQOI-RGH-----RPKDGMPVPSTAQRSSTLHTEYIVWDPQLTP-OPP 411
DB 277 DGTAGWLNGLVAERHFWREAYESGVRGVPMP-----EAFLELLWYDPRVTPAAPD 326
QY 412 ADYGTAKIHTFPNNGVVTYAGLNTQNTNFVSFKSGKLGGRAYVDVHFQPYSWIDGWR 471
DB 327 REPCTA---YFPDLGQITARTGWSAA--TCVSFKAAPGGGHRADWDEGH--RLKAAGWD 379
QY 472 SFNPGHHPQNSPTFAPNGQVVFSEALYG--PKLSHLNVLVFPAPSPSQCNKVPWEGQL 529
DB 380 AMSAGHHHPDAGAFVLHSHGAFVLADEGYSNHHKRAAHNLV-----DGEQWA 428
QY 530 GECQWLKWTG--REVGDAGEIITASQHGEMVVFSGEAYSAYSAMBLKSIVYBALLLN 587
DB 429 DE-GRYHVYEGIPERRARVDVL--AQDG-FAHATAESAMFSERLGVQKVRDLTVVTP 484
QY 588 SQTLLVVDHIERQ 600
DB 485 LGRVVIDELEAE 497

RESULT 12
Q8UBJ1 PRELIMINARY; PRT; 776 AA.
ID Q8UBJ1 AC Q8UBJ1; Q7CREI;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Oligo alginate lyase (AGR L3558P)
GN OrderedlocusNames=AGR_L3558, AtU3025;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley E., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurrello B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmieu K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollman C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
```

```
DR EMBL; AE009232; AAL43841.1; -.
DR EMBL; AE008381; AAK90358.1; -.
DR PIR; AC2928; AC2928.
DR PIR; D98354; D98354.
DR GO; GO:0016829; F:lyase activity; IEA.
KW Lyase; Complete proteome.
SQ SEQUENCE 776 AA; 87870 MW; F5889105922563E3 CRC64;

Query Match 2.9%; Score 192; DB 2; Length 776;
Best Local Similarity 19.1%; Pred. No. 8e-05;
Matches 137; Conservative 94; Mismatches 279; Indels 208; Gaps 31;

QY 33 EESVSYSEWAVTDDIDQPKTKQVDFRPNQKLKQKML-----HPSLIFDA 79
DB 99 QKSATAHSNMT-----VRSFEISEALPKTLPGRSARHAAAQTSHPRLMLNS 146
QY 80 GETQMRQ--KSRASHLHLFRAISAVTVMLSNPTYYLPP--PKHADFAAKWNIYGNL 135
DB 147 EQUSAFADAVAKDPNHCWMAEFYEKSEVPWLERPVMPPEQYPNTRVATLWRMY 202
QY 136 PPLALYCLLCPEDKVAPEFVLEYMDRMVGYKD-----MLV-----ENAPGDE 177
DB 203 -----IDQOEVIYAIRH--LATAGRVLGRDLDLDSRKWLLA VAWMDTKGATSRAYNDE 254
QY 178 VPIGHSITGPATAPDFLYNLL--DNHR--ROKYLEKIWITEEMEYKSVKSWGKQLLHN 233
DB 255 AGF-RVVVALAWGYDMLYDHLSEDERRTVRSVLLERTREVADHVIARIAHVFP---YDS 310
QY 234 HQATNMIALTGALVTGVDGSKANIWKQAVDVMEKTMFLNHI VDGSLDE-----G 286
DB 311 HAVRSLSAVLTPACIALQGESDEAGELDYTFEFL-ATLYSPWAGTDGGWAEGPHYWMWG 369
QY 287 VAYGSYTAKSVTQV--FLAQRHFNINNDNNLWVHFYFYATLPGCFORTVGIADSNY 344
DB 370 MAYLIEAANLIRSYIGYDLYQRPF-----FQNTGRFPLYTKAPGTRRAN- 413
QY 345 NWFYGPESQLVFLDKF-----ILKNAGNWLAAQRIKHPKDGMPVPSTAQRS 393
DB 414 ---FGDDSTLGLDPLGLKLYNVRQFAGVTGNGHYQWYFDHIK-----ADATGTE 459
QY 394 TLHTEYIWD-----POLTQPPADYGTAKIHTFPNNGVVTYAGLNTQNT 441
DB 460 MAFYNYGMDLNFDDLYVRHDYPQVEAVSPADLPALAVFDDIGWATIQOMEDFRLHQL- 518
QY 442 FVSFKSGKLGRAVYDVIHFQPYSWIDGWSFNPGHEHPQNSPTFAPNGVVFSEALYG 501
DB 519 FV-FKS-----SPVGSF-----SHSHGQNAFVLYAHGEDLAIQSGY- 554
QY 502 PKLSHLNVLVFPAPSPSQCNKPMWQ-----LGECAQWLKWTGEEVGDAAGEIITA 553
DB 555 -----YVAF---NSQHLNWRRTSRKNAVLIGKGQYAEKDKALARRAAGRIVSV 602
QY 554 SQHGEMVVFSGEAVSAYSSMRL-KSVYRALLNLSQTLVVDHIERQE----- 601
DB 603 EEOFGHVRIVGDATAAYQAVNPLVQKVLRETHFVNDYSYFVIVDEVCESEPOELMWLCHTL 662
QY 602 DSPINSVSAPFHNLIDIDFKIPIKFMNRYNGAMMDVMDAHYKMFVDFHHGNSPMASIQ 659
DB 663 GAPQTGRSSF-----RYNGRKA-----GPHYQVYSSGGTPOISAVE 699

RESULT 13
Q9EQCO PRELIMINARY; PRT; 411 AA.
ID Q9EQCO AC Q9EQCO;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Galactose 6-O-sulfotransferase GST-1 (Carbohydrate (Keratan sulfate
DE Gal-6) sulfotransferase 1).
DE Name=Chst1; Synonyms=Gst1;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCEI TaxID=10090;

[1] SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RX MEDLINE=21096027; PubMed=11181564; DOI=10.1093/glycob/11.1.175;
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 Rosen S.D.;
 RT "Chromosomal localization and genomic organization of the galactose/N-
 acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene
 family";
 RT Glycobiology 11:75-87 (2001).
 RL Glycobiology 11:75-87 (2001).
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Ueding T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Viallano D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Fahey J.J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RC Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RA EMBL: AF280087; AAG48245.1; -;
 DR EMBL: BC030667; AAH30667.1; -;
 DR MGD; MGI:1924219; Chet1.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransferase_1; 1.
 KW Transferase.
 SQ SEQUENCE 411 AA; 46903 MW; B1AE590EF5B9CBDC CRC64;
 Query Match 2.3%; Score 150; DB 2; Length 411;
 Best Local Similarity 18.0%; Pred. No. 0.036;
 Matches 88; Conservative 56; Mismatches 154; Indels 190; Gaps 14;
 QY 813 WILIIALWFTEL---LDVWSTCSQPICAKWTRTEAGSKLSSEGHMDL-----P 863
 Db 5 WKAVLLALASTAIQVTAIRTAFTAKSFHTCGLTDT---GLAERLCBEGPTFSVNLRSKT 61
 QY 864 DVVITLSPGSGAIIKQLFPNDSDFLYIRVPTAYIDIPETELEIDSFVDACEWVKVSDIRS 923
 Db 62 HVLILATRRSGSFGQLFNQHMDFVYLFEPL----- 93
 QY 924 GHFRLRLGWSLVQDTKLHLQNIHLHFNKGLAQYFAMNKKRKFKRRESLPEQRSQ 993
 Db 94 -----YHVQNTLIPRTQ-----KSPADRRVNLGASRDL 123
 QY 984 MKGAFDRDAEYI-----RALRHLVYPSARPV-----LSLSSGSWTL 1021
 Db 124 LRSLYDCDLYFLENIKPPVNHNTNVRFGASRVLCSPVCDPPGSSDLILEEGDCVR 193
 QY 1022 KLHFFQEVITGA-----SMRALYIVRDPRAWIYSM 1050
 Db 184 MCGLLNLTAAACRERSHVAIKTVRVPVNDLRAIVDPRLNLKVQLVDRDPRGILASR 243

Qy	1022	KUHFQEVLGA-----SMRALYIVRDPRAWIYSM	1050
D5	184	MCGLLNLTAAEACRERSHVAIKTVRVPENDLRALVEDPRLNLKVIQLVRPFGILASR	243

[illegible]

Search completed: June 23, 2005, 08:49:05
Job time : 138.102 secs

Qy 233 NHOATNMIALTGALVTCVKDKSKANIWKQAVVDVMEKTMFLNHHIVDGLSDEGVAYGSY 292
Db 1213 -----GLYCTGYQG-----EDTLVMEFYSQSS-----YSSY 1239
Qy 293 T---AKSVTQVFLAQHFNNLD--NNWLKMHFFYYATLLPGFORTVGIADSNYNWF 347
Db 1240 TDNNAPVTGLYIFADMSSDNMTAQTNYWNSY-----PQEDTVNADPDS----- 1286
Qy 348 YGPESQLVFLDKFLKNGAGNWLAAQIKRHPKDGPMVPSQAQWSTLHTEYIWDYDOLT 407
Db 1287 -----NKKVITRRVNNRYAEDYE-----IPSSV-----TSNSNYSWGDHSLT 1323
Qy 408 POPPADYGTAKIHFTFPMWVVTYAGLPN-----TQNTFVS-FKSGKLGRAV 455
Db 1324 -----MLYGSVPNITFESAEDLRLSTNMALSIHNGYAGTRRI 1363
Qy 456 YDIVHFPYSWIDGWRSFNPGHEHPDQNSFTFAPNGQVVFSEALYGPKLSHLNNVLVPAP 515
Db 1364 QCNLMKQYASLGDKFIITYDSSF--DANRFNLVP--LF----KFGKDENSDDSI CIYNE 1414
Qy 516 SPSSQCNKPHQEGQGECAQMLKWTGE-----EVGDAAGEIITASQHGEMVVFVSGEAVSAYS 571
Db 1415 NPSSE-DKRWYFSSKDDNKTDYNGGTQCIDAGTSNKDFYNNLOEIEVISVTCGYWSSYK 1473
Qy 572 SAMRLKSVYRALLLLNSQTLVVDHIERQEDSPINSVSAFFHNLDIDFKYIPYKFMNRYN 631
Db 1474 IS-----NPINI-----NTGIDSAKV--KVTVKAG 1496
Qy 632 GAMMDVNDHYKMFVPHHGNSPMASIOEABEAKKRWTFQVNVTFQWQEST--ITRIA 689
Db 1497 G-----DDQIFTADNSTVVPQ-----QPAPSFEEIYQFNNTIDCKNLNFIIDNOA 1542
Qy 690 YVFGPYINVSSCFIDSSNPGLOISLVNNTTEHVSIIVTDYHNL-----KTRFNYL 741
Db 1543 HIEIDFTATQDGRFLGAETFIIPVTCKVLGTENVIALYSENNGVQVMQIGAYRTRLNLT 1602
Qy 742 GFGGFASVADQG-----QITRFLGTQAIKVPVRHDIRIIFPFGFENIAVGLI 789
Db 1603 FAQQLVSRANGIDAVLSMETQNIQEPQLGAGTYVQLVLDKYDESHGNTKSPAIHYVDI 1662
Qy 790 LCISLVILTQWRFYLSFRKLMRWILILVIALWFIELL-----DVMSTCSQPICAKW---- 841
Db 1663 FKENDSEVIYQELSETSQTWK-----VFLSYFIEATGNKNHLM-----VRAKYQKET 1711
Qy 842 -----TRTEAE-----GSKKSLSEGHMDLPDVVITSLPGSCA--- 875
Db 1712 TDKILFORTDEKOPHGWFLSDHDKTFSGLSSAQAALKNDSEPMDF-----SGANAL 1761
Qy 876 -----ETLKOLFNSD--FLYIRVPTAYIDIPETELEIDSPVDACEWKV 918
Db 1762 YFWELFYTPMMAHRLILOQNFDAANHFWRYVNSPSGYI-----VDGKIAIYHNV 1813
Qy 919 SDIRSGHRLRLGWQLSVQDTKLHLQNIHLHPN-----RGKLAQYFA-----MN 964
Db 1814 -----RPLEEDTSNAQQLDSTDPDAVAQDDPMHYKVATFMTATLDDLMA 1858
Qy 965 KDKKXKFKRRESLPEQRSQMKGAFDRDAEYIRALR-----RHLVYPSARPVLSSSGS 1018
Db 1859 GDAAYRQLERDTLAEAKW-----YTOALNLGDEPQVMLSTTWNAPTGLNAASK 1908
Qy 1019 WTLKLFHFQEVLGASMEALYIVRPRAWIYSMLYNSKPSLYSLKNVPEHLAKLFKEGGK 1078
Db 1909 TQOQVR--QOVL-TQLRNSRVKTP-----LLGTANSLTAL-FIPQENSKL-----K 1951
Qy 1079 G-----KCNLNSGYAFYEPRLKELSKSKNAVSLLSHLWLANTAAALRINTDOLPT 1130
Db 1952 GYWRTLAORMFNLRHNLSDQPLSLPYAKPADPKALLSAAVSAGGGA----- 2001
Qy 1131 SYQLVKFEDIVH-FPQ 1145
Db 2002 --DLPKAPLTHRFPQ 2015